Efficient Methods to Reduce the Complexity of the Charge Density within Density Functional Theory for Large Systems

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Dorothea Golze

aus

Deutschland

Promotionskomitee

Prof. Dr. Jürg Hutter (Vorsitz und Leitung)

Prof. Dr. Peter Hamm

Dr. Marcella Mauri-Iannuzzi

Zürich, 2016

Abstract

Density functional theory (DFT) has become the most popular method for the quantum mechanical simulation of molecules and condensed phase systems. Even though DFT provides an excellent compromise between accuracy and performance, the simulation of large systems is still challenging in terms of computational cost. The objective of this thesis has been the development of approximate DFT-based methods that are computationally less expensive. In order to reduce the complexity of the DFT charge density, two strategies are pursued. One strategy combines different levels of theory into hybrid schemes, keeping the higher accuracy of DFT only for selected parts of the investigated system. The second strategy involves the linearization of the representation of the charge density using a resolution of identity approach.

A hybrid quantum mechanics/molecular mechanics (QM/MM) approach has been developed that includes polarization effects explicitly for large-scale simulations of adsorbate-metal systems. The adsorbate is treated by DFT and the metallic substrate is described at the MM level. The image charge formalism is employed in a self-consistent fashion in order to describe the induced charge density in the metal and the electrostatic response of the QM subsystem.

A similar QM/MM setup has been used to study the wetting behavior of water on corrugated and flat hexagonal boron nitride monolayers on Rh(111). Induction effects are in this case negligible due to the polarity of the substrate. The electrostatic properties of the substrate have been reproduced by assigning partial charges to each MM atom. A new method is introduced to determine such charges for periodic systems from restrained electrostatic potential fitting. The proposed method is especially suited to tune charges for nano-structured materials exposing surfaces and porous structures.

By introducing a local resolution of identity (LRI) approach within the Gaussian and plane waves method, the computational effort for calculating the density-dependent terms of the Kohn-Sham matrix is reduced. Significant speed-ups of the self-consistent field iterations have been observed for periodic systems. In order to reduce the LRI overhead, which becomes predominant for molecular dynamics simulations, a more efficient scheme has been developed to compute analytic integrals of contracted spherical Gaussian functions. Validation tests on a large variety of molecular systems demonstrate that errors introduced by the LRI approximation are marginal, in particular if compared to the relatively crude approximations introduced in QM/MM setups.

Zusammenfassung

Dichtefunktionaltheorie (DFT) ist mittlerweile die am häufigsten genutzte Methode um Moleküle und kondensierte Materie auf quantenmechanischer Basis zu simulieren. Obwohl DFT einen exzellenten Kompromiss zwischen Genauigkeit und Rechenaufwand bietet, ist letzterer für große Systeme immer noch zu hoch. Die Zielsetzung dieser Doktorarbeit war die Entwicklung DFT-basierter Näherungsmethoden, die weniger rechenintensiv sind. Um die Komplexität der DFT Ladungsdichte zu reduzieren, wurden zwei Ansätze verfolgt. Der erste Ansatz zielt auf die Verwendung von Hybridmodellen, wobei nur ein kleiner Teil des untersuchten Systems akkurat mit DFT beschrieben wird und der andere mit einer weniger genauen Methode. Der zweite Ansatz umfasst die Darstellung der Ladungsdichte in linearisierter Form durch Entwicklung in eine andere Basis.

Ein sogenannter QM/MM Hybridansatz, der Quantenmechanik mit klassischer Mechanik verbindet und zudem Polarisationseffekte berücksichtigt, wurde für die Simulation von Adsorbaten auf Metallen entwickelt. Die adsorbierten Moleküle werden hierbei mit DFT beschrieben und das metallische Substrat mit einem klassischen Kraftfeld. Die induzierte Ladungsdichte im Metall und die elektrostatische Reaktion des QM-Teils werden in sich konsistent mithilfe von Spiegelladungen beschrieben.

Ein ähnliches QM/MM-Modell wurde verwendet, um das Benetzungsverhalten von Wasser auf gewellten und planaren Bornitridmonolagen, welche auf Rh(111) adsorbiert sind, zu untersuchen. Das Substrat ist in diesem Fall polar und Induktionseffekte können vernachlässigt werden. Die elektrostatischen Eigenschaften des Substrats wurden reproduziert, indem jedes MM-Atom mit einer Partialladung versehen wurde. Eine neue Methode zur Bestimmung dieser Ladungen, die sich insbesondere für nanostrukturierte Oberflächensysteme und poröse Strukturen eignet, wird vorgestellt.

Ein Näherungsverfahren zur Darstellung der lokalen DFT-Dichte wurde mit der sogenannten “Gaussian-and-plane-waves”-Methode kombiniert. Zielsetzung war, den Rechenaufwand für die Beiträge zur Kohn-Sham-Matrix, die von der Dichte abhängen, zu reduzieren. Eine deutliche Beschleunigung des sogenannten “Self-Consistent-Field” Iterationsschritts wurde damit erreicht. Die lokale Näherung der Dichte ist jedoch mit einem Mehraufwand u.a. in Bezug auf die Integralberechnung verbunden, der insbesondere für Molekulardynamik-Simulationen schnell dominant wird. Deshalb wurde ein effizienteres Verfahren zur Berechnung analytischer Integrale entwickelt. Die approximative Darstellung der lokalen Dichte verursacht nur geringfügige Fehler, wie durch umfangreiche Tests gezeigt werden konnte.

Acknowledgements

First of all, I want to thank Prof. Jürg Hutter for giving me the opportunity to do my PhD in his group. I am grateful for his advice, guidance and continuous support over the last four years.

I also owe my thanks to Dr. Marcella Iannuzzi for her support and encouragement. I want to thank her for answering many of my questions covering theoretical and practical aspects of my research. I am also grateful for the many discussions and suggestions that provided valuable input for my work and for her careful revision of my manuscripts.

I thank my colleague from the Mathematics Institute, Dr. Niels Benedikter, for helping me to derive the correct expressions for the analytic integral scheme developed in this work, namely, for the integral formulas derived in Section 6.2.3. The proof of these formulas would not have been possible without his help.

I would like to express my gratitude to all members of the Hutter group for the friendly working atmosphere, academic and non-academic discussions and their support in general.

My thanks go to the Swiss National Science Foundation for funding and the Swiss National Supercomputing Center for providing generous computer resources.

Finally, I would like to thank my entire family, all friends and especially my partner Robert Richter for their support.

Publications

1. Dorothea Golze, Marcella Iannuzzi, Manh-Thuong Nguyen, Daniele Passerone and Jürg Hutter. "Simulation of Adsorption Processes at Metallic Interfaces: An Image Charge Augmented QM/MM Approach" *J. Chem. Theory Comput.* **2013**, *9*, 5086-5097

2. Dorothea Golze, Jürg Hutter and Marcella Iannuzzi. "Wetting of water on hexagonal boron nitride@Rh(111): A QM/MM model based on atomic charges derived for nano-structured substrates" *Phys. Chem. Chem. Phys.* **2015**, *17*, 14307-14316

3. Dorothea Golze, Marcella Iannuzzi and Jürg Hutter. "Local density fitting within a Gaussian and plane waves approach", in preparation

4. Dorothea Golze, Niels Benedikter, Marcella Iannuzzi and Jürg Hutter. "Integral evaluation of contracted solid harmonic Gaussian functions involving position-dependent operators", in preparation

Contents

1. Introduction

Over the last decades, computer simulations at the nanometer scale became increasingly important to study molecular structures, liquids and solids as well as interfaces between different phases. At the atomic and molecular level, quantum effects have to be considered to gain information about the electronic structure. Since chemical interactions and reactions involve changes in the electronic distribution, the accurate prediction of the latter is needed. The theoretical basis is provided by quantum mechanics which postulates the existence of a wavefunction that completely describes the quantum system. The wavefunction is obtained by solving the Schrödinger equation. However, its exact solution is practically impossible for many-electron systems, due to rapidly increasing complexity with respect to system size. Thus, approximations must be introduced. One possible ansatz is to directly use the wavefunction as basic variable. Another approach is employed in density functional theory (DFT), where the electronic density is used as variable instead. Considering an N-electron system, the wavefunction has $3N$ degrees of freedom, whereas the electronic density is defined by three spatial coordinates. Consequently, DFT is computationally much less expensive than wavefunction-based methods.

DFT became increasingly popular after the introduction of the Kohn-Sham (KS) approach which partially resolved the problem of insufficient accuracy in the functional description of the electronic structure. KS-DFT provides a fair compromise between accuracy and efficiency enabling the simulation of systems of several hundreds of atoms. However, the timescales accessible in molecular dynamics (MD) simulations are still in the range of picoseconds. Even static calculations of large condensed matter systems, metals in particular, remain challenging. To make such simulations and calculations affordable, further approximations have to be introduced. The DFT approach can be simplified by dividing the system in two parts. The smaller subsystem that is chemically of major interest, is treated by quantum mechanics (QM) and the other subsystem is described at the molecular mechanics (MM) level of theory. Such a hybrid QM/MM approach has first been proposed by Warshel and Levitt [1] in 1976 and worked out in detail by Karplus *et al.* [2] in 1990. The strength of QM/MM is that it combines the accuracy of QM methods with computationally inexpensive classical force fields. There has been a steadily increasing interest in application and development of such hybrid methods. The number of articles and reviews has increased fourfold within in the last 15 years as displayed in Figure 1.1. The importance of QM/MM has been also emphasized by awarding the 2013 Nobel Prize to Warshel,

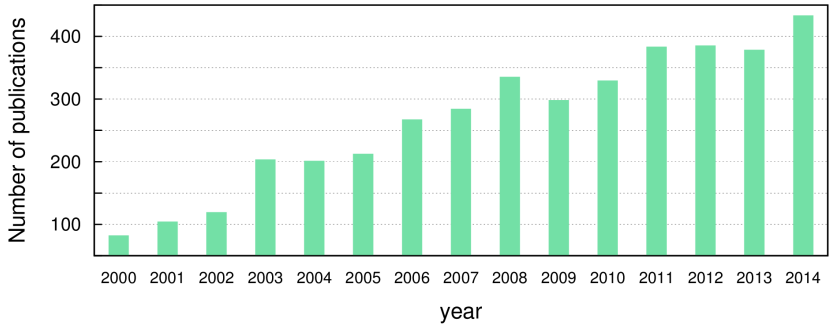

Figure 1.1.: Number of articles and reviews per year where QM/MM has been applied or QM/MM methods were developed.

Levitt and Karplus for "the development of multi-scale models for complex chemical systems". The further development of QM/MM methods focuses mainly on a better description of the MM-based interactions between QM and MM part. In particular, the electrostatic coupling between the two subsystems can be treated at different levels of sophistication and has been also subject to further improvement in this work.

Resolution of identity (RI) approximations of the total charge density are another possibility to improve the performance of DFT. The KS-DFT density is represented by a sum over the squares of one-electron orbitals, which are typically expanded in n atom-centered basis functions. In the basis set representation, the density is expressed in terms of the density matrix containing the expansion coefficients and products of the basis functions. However, this leads to $O(n^2)$ contributions of the density and an unfavorable $O(n^4)$ scaling for the term describing the Coulomb interactions between the electrons. Different density fitting techniques, also referred to as RI, have been developed in order to improve the scaling behavior and to reduce the prefactor. RI approximations can be considered as state of the art and are implemented in almost all important quantum chemistry codes. The quest for better RI techniques is still on-going and an active field of research.

The focus of this thesis has been the development of approximate DFT methods. In particular, QM/MM approaches for solid-liquid interfaces have been developed. Another key aspect is the combination of an RI approach with the Gaussian and plane waves (GPW) method. This thesis is organized in seven chapters. In the next chapter, the theoretical framework of DFT, density fitting and QM/MM is presented. In Chapter 3, an image charge augmented QM/MM approach is introduced which allows to include polarization in the simulation of adsorbate-metal systems. The electrostatic interactions between molecule (QM) and metal (MM) are accounted for by employing a self-consistent image charge approach. In Chapter 4, the wetting of water on corrugated and flat hexagonal boron nitride monolayers at Rh(111) is investigated employing a similar QM/MM setup. The electrostatic properties of the substrate are reproduced by partial charges, which are obtained from restrained electrostatic

potential fitting designed for nano-structured systems. The focus of Chapter 5 is on the development of a local RI approach within a GPW scheme. In Chapter 6, an integral scheme based on solid harmonic Gaussian functions is derived with regard to the integral types required for local RI. In Chapter 7, the results and achievements of this thesis are summarized and possible following developments are discussed.

2. Theoretical approaches

In this chapter, the theoretical framework fundamental to the methods developed in this thesis are briefly presented. In particular, density functional theory and hybrid approaches combining different levels of theory are described in general terms. To put Chapter 5 of this thesis in perspective, an overview of density fitting techniques is given.

2.1. Density functional theory

2.1.1. Key aspects

Density functional theory (DFT) has become a standard technique for solving electronic structure problems. It is broadly applied to molecular as well as condensed phase systems. Compared to wavefunction based methods, the computational cost of DFT is much lower while the accuracy is still sufficient for a wide range of systems. DFT has also been extensively used in this thesis and all method developments were embedded in a DFT framework.

Unlike in wavefunction theories, the central quantity is not the N-electron wavefunction $\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$, but the total electron density $\rho(\mathbf{r})$ defined as [3, 4]

$$\rho(\mathbf{r}) = N \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) d\omega_1 d\mathbf{x}_2 ... d\mathbf{x}_N, \tag{2.1}$$

where $\mathbf{x}_i$ are the spin orbitals. Note that we integrate over all N spin variables ω, but only over $N-1$ spatial variables $\mathbf{r}$. The ground state density ρ_0 is obtained, when Ψ corresponds to the ground state wavefunction $\Psi_0(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$.

The basic premise of DFT is that the ground state electronic energy E_0 is completely determined by ρ_0. This statement and further implications are commonly referred to as Hohenberg-Kohn theorems [5]. These theorems also imply that the ground state energy can be obtained variationally. For any trial density, E_0 is related to the energy functional $E[\rho]$ by

$$E[\rho] \geq E_0. \tag{2.2}$$

This is equivalent to saying that the density which minimizes the total energy is the exact ground state density.

In accordance with the Hohenberg-Kohn theorems, the DFT energy functional can be expressed as

$$E[\rho] = T[\rho] + V_{\text{ext}}[\rho] + V_{\text{ee}}[\rho], \tag{2.3}$$

where $T[\rho]$ is the kinetic energy of the electrons and $V_{\text{ext}}[\rho]$ represents the Coulomb interaction with an external potential usually due to the nuclei. The repulsive electron-electron interactions are summarized in $V_{\text{ee}}[\rho]$, which includes the classical Coulomb interaction and non-classical terms. Unfortunately, explicit expressions of $T[\rho]$ and $V_{\text{ee}}[\rho]$ are not known and suitable approximations have to be introduced. Early attempts to derive functionals mainly failed due to the bad representation of the kinetic energy. The most popular of these so-called orbital-free approaches is the Thomas-Fermi model [6, 7] where the derivation of $T[\rho]$ is based on the theory of the non-interacting uniform electron gas. However, also the performance of the latter is poor and the construction of accurate models for $T[\rho]$ turned out to be difficult [4, 8, 9].

2.1.2. Kohn-Sham approach

In contrast to orbital-free methods, the Kohn-Sham (KS) approach [10] allows for an accurate treatment of the kinetic energy and represented a breakthrough in DFT. By introducing an orbital basis, $T[\rho]$ can be partitioned in a non-correlated term that can be evaluated exactly and a small contribution accounting for the correlation. The general idea is to introduce a reference system of N non-interacting electrons, where the density $\rho(\mathbf{r})$ is the same as for the interacting system. In the KS approach, $\rho(\mathbf{r})$ is represented in terms of N one-electron orbitals,

$$\rho(\mathbf{r}) = \sum_{i}^{N} |\phi_i(\mathbf{r})|^2. \tag{2.4}$$

The kinetic energy functional $T_{\text{S}}[\rho]$ for the non-interacting system can be also formulated in terms of the set of orbitals $\{\phi_i\}$,

$$T_{\text{S}}[\rho] = \sum_{i}^{N} \left\langle \phi_i \left| -\frac{1}{2}\nabla^2 \right| \phi_i \right\rangle. \tag{2.5}$$

For the interacting system, the so-called exchange-correlation functional $E_{\text{xc}}[\rho]$ is introduced that contains the missing correlation in the kinetic energy and the non-classical part of the electron-electron interaction,

$$E_{\text{xc}}[\rho] = (T[\rho] - T_{\text{S}}[\rho]) + (V_{\text{ee}}[\rho] - J[\rho]). \tag{2.6}$$

The classical part of the electron-electron interaction is the Coulomb energy $J[\rho]$ given by

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \tag{2.7}$$

The general expression for the energy functional within KS-DFT can be now written as

$$E[\rho] = T_{\mathrm{S}}[\rho] + V_{\mathrm{ext}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]. \tag{2.8}$$

Equation 2.8 can be solved by applying the variational principle with respect to the orbitals. Considering additionally the orthonormality constraint, the KS equations are obtained

$$\hat{H}_{\mathrm{KS}}\phi_i = \epsilon_i\phi_i, \tag{2.9}$$

where the KS operator is given by

$$\hat{H}_{\mathrm{KS}} = \left[-\frac{1}{2}\nabla^2 + V_{\mathrm{KS}}(\mathbf{r})\right] \tag{2.10}$$

and the KS potential is

$$V_{\mathrm{KS}}(\mathbf{r}) = \frac{\delta V_{\mathrm{ext}}[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta J[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})}. \tag{2.11}$$

This set of equations can be solved iteratively by a self-consistent field procedure.

The KS formalism is in principle exact since all correlation effects are incorporated in E_{xc}. However, the explicit form of the latter is unknown and sufficient approximations have to be introduced.

2.1.3. Exchange-correlation functionals

The accuracy of a KS-DFT calculation depends heavily on the choice of the exchange-correlation (XC) functional. A general scheme for a systematic improvement of the approximation of $E_{\mathrm{xc}}[\rho]$ does not exist and the development of more accurate functionals is still on-going. Nevertheless, Perdew et *al.* suggested a classification of XC functionals referred to as "Jacob's ladder" [11]. This ladder has five rungs where for each rung more descriptors of the electronic system are introduced. For each step up the ladder, an improved accuracy can be expected, but not guaranteed.

The ladder's first rung, local density approximation (LDA) [10], includes only the electronic density as variable. In LDA, the density is treated locally and relies on the model for a uniform electron gas. The general form of the LDA functional is [4]

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho] = \int \rho(\mathbf{r})\epsilon_{\mathrm{xc}}^{\mathrm{LDA}}[\rho(\mathbf{r})]d\mathbf{r}. \tag{2.12}$$

The exchange and correlation part are treated separately and combined linearly,

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho] = E_{\mathrm{x}}^{\mathrm{LDA}}[\rho] + E_{\mathrm{c}}^{\mathrm{LDA}}[\rho] = \int \rho(\mathbf{r})\epsilon_{\mathrm{x}}^{\mathrm{LDA}}[\rho(\mathbf{r})]d\mathbf{r} + \int \rho(\mathbf{r})\epsilon_{\mathrm{c}}^{\mathrm{LDA}}[\rho(\mathbf{r})]d\mathbf{r}. \tag{2.13}$$

The exchange energy for a uniform electron gas takes a simple analytic form derived from the Dirac formula [4],

$$\epsilon_{\mathrm{x}}^{\mathrm{LDA}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}(\mathbf{r}). \tag{2.14}$$

Analytic expressions of the LDA correlation energy are only available in the limiting cases of extremely high or low densities. For densities in between, interpolating formulas have been constructed by, e.g., Vosko, Wilk and Nusair (VWN) [12] or Perdew and Wang (PW) [13]. LDA is mainly used in solid state physics where the approximation of a uniform electron gas is often sufficient. For molecular systems, LDA underestimates the exchange and overestimates the correlation energy and predicts bond strengths that are usually 100 kJ/mol too large [4, 14].

The next step on the Jacob's ladder is the generalized gradient approximation (GGA) [15–17], which includes additionally the gradient of the density,

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int \rho(\mathbf{r})\epsilon_{\mathrm{xc}}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r})]d\mathbf{r}. \tag{2.15}$$

Various GGA functionals have been proposed. For molecular systems, one of the most popular ones is the BLYP functional, which is a combination of the Becke 1988 exchange functional [18] and the Lee-Yang-Parr (LYP) correlation functional [19]. The Perdew-Burke-Ernzerhof (PBE) functional [20] is usually favored for materials, but performs also well for molecular systems. The exchange part of PBE is given in terms of the LDA exchange,

$$\epsilon_{\mathrm{x}}^{\mathrm{PBE}} = \epsilon_{\mathrm{x}}^{\mathrm{LDA}} F(s), \tag{2.16}$$

where $F(s)$ can be considered as an enhancement factor defined as

$$F(s) = 1 + \alpha - \frac{\alpha}{1 + \beta s^2}. \tag{2.17}$$

This factor depends on the dimensionless gradient variable s given by

$$s = \frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})}. \tag{2.18}$$

The correlation part is also related to its LDA counterpart by adding a function $H(t)$, where the variable t depends on s as well,

$$\epsilon_{\mathrm{c}}^{\mathrm{PBE}} = \epsilon_{\mathrm{c}}^{\mathrm{LDA}} + H(t). \tag{2.19}$$

The explicit form of $H(t)$ can be found in Refs. [4, 20]. The parameters in the PBE functional like α or β do not depend on experimental input, but are derived from constraints that the exact XC functional theoretically should satisfy. These are, for example, conditions concerning the asymptotic behavior of the exchange and correlation potential or an upper bound for the XC energy with respect to LDA as suggested by Lieb and Oxford [21]. Since the method development in this thesis focused on condensed phase systems, PBE has been predominately used for the associated applications.

The third rung of the ladder are the meta-GGA methods. These methods include the orbital kinetic density τ [4],

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_{i}^{occ} |\nabla\phi_i(\mathbf{r})|^2, \tag{2.20}$$

where *occ* indicates that the sum is running over the number of occupied orbitals. A commonly used meta-GGA functional is, e.g., TPSS (Tao-Perdew-Staroverov-Scuseria) [22] which is also employed for materials and can be considered as meta-GGA equivalent of the PBE functional. The inclusion of τ comes with almost no additional computational cost. This changes for the next rung, which involves so-called hybrid functionals. These functionals include exact Hartree-Fock exchange, which means that $E_{\mathrm{x}}^{\mathrm{DFT}}[\rho]$ is partly replaced with its orbital analogue $E_{\mathrm{x}}^{\mathrm{HF}}[\Psi]$ [4]. Including exact exchange often improves the accuracy, but increases the computational cost by at least an order of magnitude. Popular hybrid functionals are, e.g., B3LYP [23] or PBE0 [24]. The fifth rung of the ladder is the random phase approximation (RPA) [25–27]. The essence of this approximation is to employ virtual orbitals as a further variable in order to improve, among others, on dispersion interactions.

2.1.4. Dispersion correction

It is commonly known that LDA, GGA, meta-GGA and hybrid functionals fail to describe dispersion interactions [28]. The correct modeling of van der Waals (vdW) interactions is especially important for weakly interacting systems, e.g., stacked DNA base pair configurations or molecules physisorbed at metal surfaces. But also for systems with strong electrostatic interactions such as ionic liquids, dispersion interactions play a crucial role and cannot be neglected [29–32].

To incorporate vdW forces, several corrections have been proposed. A concise overview of the existing methods is given in Ref. [28]. The simplest models are the DFT-D corrections, where an empirical correction term is added to the KS-DFT energy,

$$E_{\mathrm{DFT-D}} = E_{\mathrm{DFT}} + E_{\mathrm{disp}}. \tag{2.21}$$

The long range interactions are assumed to be pairwise additive and show the correct $1/r_{\mathrm{AB}}^6$ asymptotic behavior,

$$E_{\mathrm{disp}} = -\sum_{\mathrm{AB}} \frac{C_6^{\mathrm{AB}}}{r_{\mathrm{AB}}^6}, \tag{2.22}$$

where r_{AB} is the distance between particle A and B. Faster decaying C_n/r_{AB}^n terms with $n > 6$ are usually neglected. The dispersion coefficients C_6^{AB} are specific for each element combination and are usually referenced to experimental data. The most popular DFT-D scheme is Grimme's DFT-D2 correction [33]. The C_6 coefficients for the latter are computed from ionization potentials and static polarizabilities. To avoid that E_{disp} diverges for small r_{AB}, a damping function is introduced.

One of the short-comings of these simple DFT-D corrections is that the dispersion coefficients are independent of the chemical environment. However, the C_6 coefficients can significantly vary for different oxidation or hybridization states. The DFT-D3 scheme of Grimme [34] accounts for environmental effects by treating the dispersion coefficients as function of the coordination number. Depending on the number of neighbors, the appropriate coefficient is assigned to each pair AB during each simulation step. As an additional improvement, DFT-D3 includes also the C_8/r_{AB}^8 interaction term. Other models like the scheme by Tkatchenko and Scheffler (vdW-TS) [35] or the Becke-Johnson (BJ) model [36–38] are also capable to capture the environmental dependence of the dispersion coefficients. In the vdW-TS scheme, this is achieved by considering effective atomic volumes, whereas in the BJ model the polarizabilities of the atoms are considering in a rather sophisticated way. A further problem of the empirical approaches is, that the dispersion interactions are not solely pairwise attractive, which is especially important for condensed phase systems. In DFT-D3, this is empirically accounted for by introducing an additional three-body term.

A non-empirical approach for including dispersion is provided by the so-called vdW density functionals (vdW-DFs). The vdW functionals do not rely on experimental parameters, but directly determine the dispersion contribution from the electronic density. To account for the missing long-range part in local or semi-local functionals such as LDA or GGA, a non-local correlation term is added. The general form of the non-local correlation energy is [28, 39]

$$E_{\mathrm{c}}^{\mathrm{nl}} = \iint \rho(\mathbf{r})\varphi(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')d\mathbf{r}d\mathbf{r}', \tag{2.23}$$

where $\varphi(\mathbf{r},\mathbf{r}')$ is a function depending on $\mathbf{r}-\mathbf{r}'$. Initially, vdW-DFs overestimated the dispersion interactions and were computationally more demanding than GGA or hybrid functionals. Thanks to more recent improvements on the accuracy and computational performance of vdW-DF [40–42], these type of functionals are becoming increasingly popular. However, vdW functionals are a pairwise approach as well and will fail for systems where higher order dispersion terms cannot be neglected. In the context of DFT, RPA is the most advanced method. It includes also many-body dispersion effects, but comes with a much higher computational cost.

2.2. Density fitting

2.2.1. Incentive

In Kohn-Sham DFT, the electron density ρ is represented in terms of the molecular orbitals ϕ_i, see Equation 2.4. Since the KS operator already depends on $\{\phi_i\}$, the KS equations can be only solved numerically for atomic calculations. In molecular

calculations, the KS differential equations are converted to a set of algebraic equations by expanding the orbitals into a set of atom-centered basis functions $\{\chi_\mu\}$,

$$\phi_i(\mathbf{r}) = \sum_\mu C_{\mu i}\chi_\mu(\mathbf{r}). \tag{2.24}$$

The electron density can be rewritten in terms of the density matrix $P_{\mu\nu}$ as

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}), \tag{2.25}$$

where the matrix elements $P_{\mu\nu}$ for closed-shell systems are given by

$$P_{\mu\nu} = 2\sum_i^{occ} C_{\mu i}C_{\nu i}. \tag{2.26}$$

Employing the basis set expansion, the expression for the Coulomb term defined in Equation 2.7 modifies to

$$J[\rho] = \frac{1}{2}\sum_{\mu,\nu,\kappa,\lambda} P_{\mu\nu}P_{\kappa\lambda}\iint \frac{\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})\chi_\kappa(\mathbf{r}')\chi_\lambda(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'. \tag{2.27}$$

The KS density can be further considered as a sum of pair-contributions,

$$\rho(\mathbf{r}) = \sum_{\mathrm{AB}}\rho_{\mathrm{AB}} = \sum_{\mathrm{AB}}\sum_{\mu\in\mathrm{A}\nu\in\mathrm{B}} P_{\mu\nu}\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}), \tag{2.28}$$

where ρ_{AB} is the density obtained for atom pair AB. Inserting Equation 2.28 in the Coulomb term, yields the four-center two-electron repulsion integrals (ERIs),

$$(\mu_{\mathrm{A}}\nu_{\mathrm{B}}|\kappa_{\mathrm{C}}\lambda_{\mathrm{D}}) = \iint \frac{\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})\chi_\kappa^{\mathrm{C}}(\mathbf{r}')\chi_\lambda^{\mathrm{D}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'. \tag{2.29}$$

The evaluation of these integrals scales with n^4, where n is the number of basis functions. Thus, the computation of the Coulomb term will dominate the calculation and limit the system size. Additionally, the expansion of the density leads to n^2 contributions for $E_{\mathrm{xc}}[\rho]$ and $V_{\mathrm{ext}}[\rho]$.

To improve the scaling especially of the Coulomb term, many resolution of identity (RI) approaches for representing the density have been developed. Density fitting techniques can be roughly divided in approximations of

1. each product $\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})$
2. each pair density ρ_{AB} (local fitting)
3. full density $\rho(\mathbf{r})$ (global fitting).

Other approaches rely on a projection and cubic spline expansion of the density [43] or an alternative representation in, e.g., plane waves [44]. However, these methods are not considered in the following. Density fitting techniques have been implemented for KS-DFT [45–47], but also within Hartree-Fock [48–50], second-order Møller-Plesset (MP2) [51–54] and coupled cluster frameworks [55, 56].

2.2.2. Pair-product fitting

The first attempts to reduce the scaling of $J[\rho]$ date back to the 1940s. One of the first approaches is the Zero-Differential-Overlap (ZDO) scheme [57, 58] where multi-center integrals are completely neglected. The product of two different orbitals is set to zero,

$$\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}) \approx \delta_{\mu\nu}\delta_{\mathrm{AB}}\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}) \tag{2.30}$$

and the remaining integrals are parametrized to compensate for this crude approximation. Different flavors of ZDO exist such as the Intermediate Neglect of Differential Overlap (INDO) [59], which keeps additionally two-center ERIs of the type $(\mu_{\mathrm{A}}\mu_{\mathrm{A}}|\nu_{\mathrm{B}}\nu_{\mathrm{B}})$. The Neglect of Diatomic Differential Overlap (NDDO) [60] represents an improvement over INDO since also the $(\mu_{\mathrm{A}}\nu_{\mathrm{A}}|\kappa_{\mathrm{B}}\lambda_{\mathrm{B}})$ integrals are retained.

The Mulliken approximation [61, 62] was proposed in 1949 and states that the product $\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})$ can be reformulated as an infinite expansion of χ_μ^{A} and χ_ν^{B} using the overlap matrix elements $S_{\mu\nu}$ as expansion coefficients. If the two centers A and B are not too far apart, the leading term in the expansion is given by

$$\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}) \approx \frac{1}{2}S_{\mu\nu}\left[\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\mu^{\mathrm{A}}(\mathbf{r}) + \chi_\nu^{\mathrm{B}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})\right]. \tag{2.31}$$

Within the Mulliken approximation, the ERIs are reduced to one and two-center integrals. The Rüdenberg approximation [63] is conceptually similar to Mulliken's ansatz, but considers cases where centers A and B are farther apart from each other. The Löwdin scheme [64] is as well similar to the Mulliken approximation, but introduces more sophisticated expansion coefficients $\lambda_{\mu\nu}$

$$\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}) \approx \lambda_{\mu\nu}^{\mathrm{A}}\left[\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\mu^{\mathrm{A}}(\mathbf{r})\right] + \lambda_{\mu\nu}^{\mathrm{B}}\left[\chi_\nu^{\mathrm{B}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})\right], \tag{2.32}$$

which also include the parameters α and β determined from conditions on total charges and moments,

$$\lambda_{\mu\nu}^{\mathrm{A}} = S_{\mu\nu}(1+\alpha/\beta)^{-1} \qquad \lambda_{\mu\nu}^{\mathrm{B}} = S_{\mu\nu}(1+\beta/\alpha)^{-1}. \tag{2.33}$$

All these early methods for approximating $\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})$ suffer from several deficiencies such as no rotational invariance or a poor representation of nodal structures. The Projection of Diatomic Differential Overlap (PDDO) [65] improves on these shortcomings by expressing the two-center product of basis functions as linear combination of all possible one-center charge distributions,

$$\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}) \approx \sum_{m\in\mathrm{A},n\in\mathrm{A}} C_{\mu\nu mn}\chi_m^{\mathrm{A}}(\mathbf{r})\chi_n^{\mathrm{A}}(\mathbf{r}) + \sum_{k\in\mathrm{B},l\in\mathrm{B}} C_{\mu\nu kl}\chi_k^{\mathrm{B}}(\mathbf{r})\chi_l^{\mathrm{B}}(\mathbf{r}). \tag{2.34}$$

The expansion is entirely expressed in terms of functions that are part of the basis set. The coefficients $C_{\mu\nu mn}$ and $C_{\mu\nu kl}$ are obtained by minimizing the error in the overlap integral. The Limited Expansion of Diatomic Overlap (LEDO) [66] uses the same expansion, but minimizing the error in the Coulomb integral.

2.2.3. Local and global fitting

In local fitting schemes, the pair densities ρ_{AB} are fitted instead of the individual products $\chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})$. This requires the introduction of an independent set of auxiliary functions. The pair densities are then approximated by an expansion in one-center auxiliary functions $\{f_i\}$,

$$\rho_{\mathrm{AB}} \approx \sum_i a_i^{\mathrm{A}} f_i^{\mathrm{A}}(\mathbf{r}) + \sum_j a_j^{\mathrm{B}} f_j^{\mathrm{B}}(\mathbf{r}). \tag{2.35}$$

The expansion coefficients $\{a_i^{\mathrm{A}}\}$ and $\{a_j^{\mathrm{B}}\}$ are obtained by a least square fitting procedure. More details on local density fitting approaches are given in Chapter 5.

Global density fitting approaches have been first introduced by Sambe and Felton [67] in 1975, were further developed by Dunlap [68] and Vahtras *et al.* [69] and became popular by the extensive work of Ahlrichs and co-workers [45, 70, 71] in the 1990s. Instead of fitting pair densities, an approximation of the full density is attempted using again an expansion in auxiliary functions,

$$\rho(\mathbf{r}) \approx \sum_i c_i f_i(\mathbf{r}) = \tilde{\rho}(\mathbf{r}). \tag{2.36}$$

The expansion coefficients $\{c_i\}$ are obtained by minimizing the error in the Coulomb self-interaction,

$$\iint \frac{[\rho(\mathbf{r}) - \tilde{\rho}(\mathbf{r})][\rho(\mathbf{r}') - \tilde{\rho}(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' = \min. \tag{2.37}$$

Global density fitting techniques are mainly used for the Coulomb part and are also widely used in MP2 implementations. Within a KS-DFT context, the application of global RI for $E_{\mathrm{xc}}[\rho]$ is rarely reported since the XC term can be efficiently calculated by numerically integration with a formally linear scaling [72, 73]. However, fully variational RI approaches have been also derived for the XC part [74].

The advantage of global and local fitting compared to methods like PDDO or LEDO is that the full or local density is only fitted once. Using PDDO or LEDO, the $n_{\mathrm{A}} n_{\mathrm{B}}$ products have to be fitted for each pair, where n_{A} and n_{B} are the basis functions at A and B, respectively. An obstacle of the global and local fitting approaches is that a set of a carefully optimized auxiliary basis functions has to be provided, whereas PDDO and LEDO do not require any further external input.

The auxiliary basis sets contain generally more functions than the primary basis sets and include functions with higher l quantum numbers. Auxiliary basis sets for global RI schemes are usually larger by a factor of two. The auxiliary basis sets for local RI have to be more extended. They are on average 3-4 times larger than the primary basis sets. The reason is that the error in the local fit is linear, whereas the error is quadratic for global schemes [68, 75]. This is due to the fact that another metric is used for the local fit, which is explained more in detail in Chapter 5. Furthermore, the presence of a third atom C providing additional basis functions cannot be exploited

since the fit is restricted to the pair AB. To compensate for this, the basis set has to be increased. On the one hand, a larger fit basis set is apparently a disadvantage and computationally more demanding. On the other hand, the pair fit is independent of its chemical environment. The quality of the fit is not affected whether AB constitutes a diatomic molecule or a pair within a larger structure.

2.3. Hybrid approaches - QM/MM

Quantum mechanical (QM) methods are able to accurately describe the electronic structure of molecular and condensed phase systems, but come with a high computational cost. Even though the computational cost of DFT is relatively low compared to wavefunction-based methods, it is still limited to a few hundred or a few thousand atoms. However, enzymes and other large biomolecules easily reach sizes of several 100 000 atoms. Furthermore, ab initio molecular dynamics simulations at the DFT level are restricted to the picosecond timescale [76]. Force-field based methods allow for simulation times in the nanosecond range and larger system sizes, but are generally less accurate and cannot describe electronic processes. QM/MM hybrid schemes [1] combine methods at different levels of theory. The chemically interesting region is treated by QM while its environment, e.g. solvent molecules, are treated by molecular mechanics (MM).

In QM/MM, the Hamiltonian is partitioned in three parts,

$$\hat{H}_{\text{total}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}}, \tag{2.38}$$

where $\hat{H}_{\text{QM}}$ and $\hat{H}_{\text{MM}}$ describe the QM and MM subsystems, respectively. The last term defines the interactions between the QM and MM system and is treated at the MM level of theory. The resulting total energy of the partitioned Hamiltonian is

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}}. \tag{2.39}$$

The definition of the QM/MM energy is not straightforward and has to be carefully chosen for each system. vdW and electrostatic interactions have to be considered. If the QM system is linked to the MM subsystem by covalent bonds, bonded terms have to be incorporated as well,

$$E_{\text{QM/MM}} = E^{\text{bond}}_{\text{QM/MM}} + E^{\text{vdW}}_{\text{QM/MM}} + E^{\text{elec}}_{\text{QM/MM}}. \tag{2.40}$$

Bonds cut by the QM/MM boundary can, e.g., occur in QM/MM setups of large biomolecules. Different schemes to cope with covalently coupled subsystems are reviewed in Ref. [77]. The conceptually simplest approach is the link-atom method [2], where the dangling bond of the QM subsystem is saturated with, e.g., a hydrogen atom (link atom). The covalent bonds crossing the boundary are modeled by classical force fields.

Dispersion interactions between QM and MM atoms are usually accounted for by the Lennard-Jones (LJ) potential [78]. The LJ parameters ϵ_{AB} and σ_{AB} are assigned to both, QM atom A and MM atom B,

$$E_{\mathrm{LJ}} = \sum_{\mathrm{AB}} 4 \cdot \epsilon_{\mathrm{AB}} \left[\left(\frac{\sigma_{\mathrm{AB}}}{r_{\mathrm{AB}}} \right)^{12} - \left(\frac{\sigma_{\mathrm{AB}}}{r_{\mathrm{AB}}} \right)^{6} \right], \tag{2.41}$$

where r_{AB} is the distance between A and B. The first term in Equation 2.41 describes Pauli repulsion, while $-1/r_{\mathrm{AB}}^6$ is the typically vdW term used in similar form for the DFT-D approach discussed above.

Electrostatic interactions can be treated at different levels of complexity and the QM/MM development in this thesis focused on improvements of these interactions. The different electrostatic embedding models include polarization effects between the QM and MM subsystem to a different extent, see reviews [77,79]. The simplest model is mechanical embedding where a set of pre-calculated partial charges is also assigned to the QM atoms. The charge-charge interactions between QM and MM atoms are simply treated by classical expressions. The major drawback of this approach is that the QM charge density does not respond to the presence of the electrostatic field generated by the MM charges. Furthermore, mechanical embedding is rarely useful in simulations. If the QM configuration changes during, e.g., a reaction, a re-evaluation of the point charges would be required.

In the most popular scheme, electrostatic embedding, the MM charges can polarize the QM region. This achieved by including the MM point charges in the QM one-electron Hamiltonian [77],

$$\hat{H}^{\mathrm{el}}_{\mathrm{QM/MM}} = -\sum_{i}^{N} \sum_{a \in \mathrm{MM}}^{M} \frac{q_a}{|\mathbf{r}_i - \mathbf{R}_a|} + \sum_{j \in \mathrm{QM}}^{L} \sum_{a \in \mathrm{MM}}^{M} \frac{q_a Z_j}{|\mathbf{R}_j - \mathbf{R}_a|}, \tag{2.42}$$

where N is the number of electrons and L and M are the number of QM and MM atoms, respectively. The set of MM points charges $\{q_a\}$ are located at $\mathbf{R}_a$ and the QM atoms with nuclear charges Z_j at the center $\mathbf{R}_j$. Using electrostatic embedding, the QM charge distribution can adapt in every step of the electronic structure optimization to the electrostatic environment. No charges have to be derived for the QM atoms and the QM charge density is coupled to the configuration of the MM system. However, the electrostatic embedding is computationally more demanding than the mechanical approach. If the number of MM atoms is rather large, the evaluation of the one-electron integrals will dominate the calculation.

A further improvement is introduced by polarized embedding methods where the QM atoms can polarize the MM regions as well. Various models for treating polarization have been developed and are exhaustively discussed in reviews [80–85]. In most schemes, the QM charge density is unaffected by polarization of the MM region. On a higher level of sophistication, QM density and MM charges mutually modify

each other in a fully self-consistent fashion. So-called Drude oscillator models [86–88] represent the induced dipoles by a massless point charge attached to the MM atom via a spring. Other approaches are based on point dipole induction models [80] or rely on distributed multipole analysis [89] of the QM part. However, most of these polarization model have been developed for biomolecular systems and their application to different types of systems is not straightforward. In Chapter 3 of this thesis, a polarized embedding method for adsorbate-metal systems is developed. On the contrary, the QM/MM setup for the wetting of the nanomesh, Chapter 4, neglects polarization and is based on electrostatic embedding discussed above.

3. Simulation of adsorption processes at metallic interfaces: an image charge augmented QM/MM approach [90]

In this chapter, a novel method for including polarization effects within hybrid quantum mechanics/molecular mechanics (QM/MM) simulations of adsorbate-metal systems is presented. The interactions between adsorbate (QM) and metallic substrate (MM) are described at the MM level of theory. Induction effects are additionally accounted for by applying the image charge formulation. The charge distribution induced within the metallic substrate is modeled by a set of Gaussian charges (image charges) centered at the metal atoms. The image charges and the electrostatic response of the QM potential are determined self-consistently by imposing the constant-potential condition within the metal. The implementation is embedded in a highly efficient Gaussian and plane wave framework and is naturally suited for periodic systems. Even though the electronic properties of the metallic substrate are not taken into account explicitly, the augmented QM/MM scheme can reproduce characteristic polarization effects of the adsorbate. The method is assessed through the investigation of structural and electronic properties of benzene, nitrobenzene, thymine and guanine on Au(111). The study of small water clusters adsorbed on Pt(111) is also reported in order to demonstrate that the approach provides a sizable correction of the MM-based interactions between adsorbate and substrate. Large-scale molecular dynamics (MD) simulations of a water film in contact with a Pt(111) surface show that the method is suitable for simulations of liquid/metal interfaces at reduced computational cost.

3.1. Introduction

Interfaces between metallic substrates and molecular adsorbates are present in electrolysis, heterogeneous catalysis, transistors, metal-supported nanodevices and many other applications. The accurate theoretical description of such interfaces requires large simulation cells consisting of several layers in the metallic substrate and sufficient lateral dimensions. A full quantum mechanical treatment of these systems, e.g. by means of density functional theory (DFT) is often not affordable due to computational limitations. Computationally less expensive approaches, like molecular mechanics (MM), can provide a solution, but cannot capture electronic effects. An

alternative strategy is the usage of a hybrid scheme combining QM and MM methods [1]. In a QM/MM setup of an adsorbate-metal system, the electronic structure of the adsorbate is of primary importance. Thus, the adsorbate is treated by QM and the metal by classical force fields. Such a model implies that the interactions between adsorbate and substrate are described by an empirical potential. However, with standard force fields it is not straightforward to reproduce polarization effects explicitly. The polarization of metal and adsorbate is caused by electrostatic screening effects in metallic conductors. The electronic distribution of the metal responds to the electrostatic potential generated by the adsorbate, such that the electrostatic potential within the metal is zero or at least a constant when an external potential is applied. This response can be modeled in terms of an image charge distribution within the metal [91, 92]. Image charges interact with the charge density of the adsorbate and can affect structural and electronic properties of the adsorbed molecule. Experimental and theoretical studies [93, 94] showed that effects associated to the induced charge distribution drive the self-assembly of organic molecules on metallic substrates. Including polarization is also crucial for modeling electrochemical interfaces at metal electrodes. Wernersson *et al.* demonstrated that neglecting image charges can lead to highly inaccurate forces on the electrolytes and qualitatively wrong results for certain system parameters [95]. Image charge interactions are also present between the tip of an atomic force microscope (AFM) and the sample generating significant forces on the tip. Therefore, taking these phenomena into account is also important in the interpretation of AFM images, as discussed in Refs. [96–98].

Several models for capturing polarization effects at metallic interfaces within MM schemes have been addressed in literature. An approach called discrete classical model (DCM) was introduced by Finnis *et al.* [99,100] and used for simulations of isopropanol on Pt(111) [101]. The basic idea behind this method is to model the polarization of the metal by a set of point charges and dipoles assigned to each metal atom. Another MM-based approach tries to reproduce the polarization of the metal by attaching two opposite charges, which are connected by a spring potential or rigid rod, to each atom in the metallic substrate [102,103]. Both methods do not account for the fact that the image charges modify the charge distribution of the adsorbate and that this in turn changes the polarization of the metal. Therefore, a variational procedure adjusting the induced charges on-the-fly is required. The Gaussian charge method developed by Siepmann and Sprik [104] treats the image charges as dynamical variables and realizes the mutual modification of induced and adsorbate charges. The Siepmann-Sprik scheme has been successfully applied to water/platinum interfaces [104–106] as well as ionic liquids at platinum [107] and carbon [108,109] electrodes within a purely classical frame.

Even though QM/MM frameworks are very popular for simulations of biochemical systems [77, 110–113], there are only a few cases where QM/MM models have been applied to metallic substrates [114, 115]. Self-consistent techniques are available for

the inclusion of polarization effects in QM/MM simulations of bioorganic molecules [89, 116, 117], but there is no such method, to our knowledge, available for metallic interfaces. This motivates us to adapt the Siepmann-Sprik approach to a QM/MM scheme.

Here, we present an implementation of the image charge (IC) method within a QM/MM framework, where the QM part is treated at the Kohn-Sham (KS) DFT level within a mixed Gaussian and plane wave [118] (GPW) formalism. In the next section, the theoretical setup is introduced and the implementation details are described. Efficiency and scalability of our method, which is referred to as IC-QM/MM in the following, are also discussed. Different test systems, such as small aromatic molecules adsorbed on Au(111), are used to assess the new approach. We investigate how the interactions of the DFT charge density with the image charges affect interaction and adsorption energies as well as the electronic structure of the adsorbate. The IC-QM/MM approach is applied to small water clusters on Pt(111) and the accuracy of our approach is evaluated by comparing structural and energetic properties to experimental data and full DFT results. "Full DFT" in this context means, that metal as well as adsorbate are treated at the DFT level. Furthermore, results for a large-scale MD simulation of a liquid water/Pt(111) interface are discussed and related to experimental studies.

3.2. Theory and implementation

3.2.1. QM/MM formulation of the image charge approach: IC-QM/MM

The charge distribution $\rho(\mathbf{r})$ of a molecule approaching the metallic surface generates an electrostatic potential (Hartree potential) $V_H(\mathbf{r})$ that extends beyond the surface into the substrate. $V_H(\mathbf{r})$ induces an electrostatic response of the metallic conductor such that the electrostatic potential within the metal is constant. To model this response, Siepmann and Sprik [104] introduced the induced charge distribution $\rho_m(\mathbf{r})$, which in turn generates the potential $V_m(\mathbf{r})$ screening $V_H(\mathbf{r})$ in the region of the substrate. Siepmann and Sprik proposed to define $\rho_m(\mathbf{r})$ as a linear expansion of Gaussian functions centered at the metal atoms,

$$\begin{aligned}\rho_m(\mathbf{r}) &= \sum_a c_a g_a(\mathbf{r}, \mathbf{R}_a) \\ &= \sum_a q_a \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha|\mathbf{r} - \mathbf{R}_a|^2\right),\end{aligned} \tag{3.1}$$

where $\mathbf{R}_\mathrm{a}$ is the position of metal atom a and g_a the spherical Gaussian function located at a. The width of all Gaussian functions is determined by one single, fixed

parameter α and q_a can be interpreted as magnitude of the induced charge at each position a. The expansion coefficients c_a are unknown and are obtained by imposing the condition that the electrostatic potential is constant for each point $\mathbf{r}$ within the metallic slab,

$$V_H(\mathbf{r}) + V_m(\mathbf{r}) = \int \frac{\rho(\mathbf{r}') + \rho_m(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' = V_0. \tag{3.2}$$

In this expression, V_0 is a constant potential that can be different from zero, if an external potential is applied. While in the original Siepmann-Sprik method $\rho(\mathbf{r})$ is obtained from the distribution of point charges associated with the force field, in IC-QM/MM $\rho(\mathbf{r})$ is the KS density plus the charge density assigned to the atomic cores,

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}) + \rho_{\text{core}}(\mathbf{r}). \tag{3.3}$$

$P_{\mu\nu}$ is the density matrix and $\{\chi_\mu\}$ represents the basis of contracted Gaussian functions used to expand the molecular orbitals.

The electrostatic energy E_{el} is given by

$$E_{el} = \frac{1}{2}\int\int \frac{(\rho(\mathbf{r}) + \rho_m(\mathbf{r}))(\rho(\mathbf{r}') + \rho_m(\mathbf{r}'))}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}d\mathbf{r}' - \int V_0\rho_m(\mathbf{r})d\mathbf{r}. \tag{3.4}$$

Instead of imposing the condition of Equation 3.2, this constraint is modified such that only the weighted integral of Equation 3.2 is forced to a constant. This is equivalent to requiring that the gradients of E_{el} with respect to the coefficients c_a are zero, i.e. E_{el} has to be minimized as function of the variables c_a. Thereby, the following set of linear equations is obtained,

$$\begin{aligned}
\frac{\partial E_{el}}{\partial c_a} &= 0 \\
&= \int (V_H(\mathbf{r}) + V_m(\mathbf{r}) - V_0) \cdot g_a(\mathbf{r})d\mathbf{r} \\
&= \int (V_H(\mathbf{r}) - V_0) g_a(\mathbf{r})d\mathbf{r} + \sum_b c_b \int\int \frac{g_b(\mathbf{r}') \cdot g_a(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}d\mathbf{r}' \\
&= \int (V_H(\mathbf{r}) - V_0) g_a(\mathbf{r})d\mathbf{r} + \sum_b c_b T_{ab}.
\end{aligned} \tag{3.5}$$

Here, T_{ab} are the Coulomb integrals between two spherical Gaussian functions centered at two different metallic centers. These integrals are stored in an $N_{\text{IC}} \times N_{\text{IC}}$ matrix, where N_{IC} is the number of metal centers carrying an image charge. Solving the linear set of equations yields the unknowns c_a and the potential $V_m(\mathbf{r})$ can be calculated. This potential is then added to $V_H(\mathbf{r})$ and the corrected electrostatic term is added to the Hamiltonian. New molecular orbitals and a new KS charge density

are obtained by solving the KS equations for the corrected Hamiltonian. As usual in DFT calculations, this procedure is iterated until self-consistent convergence of the electronic structure optimization is achieved. Thanks to the self-consistent optimization scheme, the polarization of the QM subsystem induced by the interaction with the metal is automatically taken into account.

It can be easily demonstrated that our QM/MM formulation of the IC method provides the correct solution for E_{el}. The electrostatic energy of a charge distribution in the vicinity of a metal is equivalent to the Coulomb energy between the real charges and *half* of the Coulomb interaction energy between real charges and image charges [91]. Assuming for simplicity that $V_0 = 0$, the electrostatic energy given in Equation 3.4 has to be consequently equivalent to:

$$E_{el} = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\cdot\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}d\mathbf{r}d\mathbf{r}' + \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\cdot\rho_m(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}d\mathbf{r}d\mathbf{r}'. \tag{3.6}$$

Employing the condition given by Equation 3.5 (2nd line), it can be shown that the other two terms indeed vanish,

$$\frac{1}{2}\int\int \rho_m(\mathbf{r})\frac{\rho(\mathbf{r}')+\rho_m(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}d\mathbf{r}d\mathbf{r}' = \frac{1}{2}\sum_a c_a \underbrace{\int g_a(\mathbf{r})(V_H(\mathbf{r})+V_m(\mathbf{r}))d\mathbf{r}}_{=0}. \tag{3.7}$$

The IC-QM/MM approach differs from the original Siepmann-Sprik method in several aspects. In the classical version of the IC approach, the potential is rigorously forced to a constant at every metal site $\mathbf{R}_a$, whereas in IC-QM/MM a condition on the potential is imposed at every real-space grid point $\mathbf{r}$. The condition we employ is that the integral of the potential weighted by a localized Gaussian function equals zero, see Equation 3.5. This guarantees that the quantity $V_H + V_m$ is approximately constant over the whole space of the metallic slab, which is not the case in the original scheme. Therefore, IC-QM/MM is expected to be numerically more stable with respect to the parameter α as discussed below.

When forces are required, e.g. for geometry optimizations or MD runs, the force contributions of the image charges have to be considered. The total gradient of the electrostatic energy with respect to the atomic positions $\mathbf{R}_i$ of the QM atoms is

$$\begin{aligned}\frac{\partial E_{el}}{\partial \mathbf{R}_i} &= \int \frac{\partial E_{el}}{\partial \rho(\mathbf{r})}\frac{\partial \rho(\mathbf{r})}{\partial \mathbf{R}_i}d\mathbf{r} \\ &= \sum_{\mu\nu} P_{\mu\nu}\int (V_H(\mathbf{r})+V_m(\mathbf{r}))\frac{\partial \chi_\mu(\mathbf{r},\mathbf{R}_i)\chi_\nu(\mathbf{r},\mathbf{R}_i)}{\partial \mathbf{R}_i}d\mathbf{r}.\end{aligned} \tag{3.8}$$

The contribution of the image charges to the gradients on the MM atoms (metal) are given by

$$\begin{aligned}\frac{\partial E_{el}}{\partial \mathbf{R}_a} &= \int \frac{\partial E_{el}}{\partial \rho_m(\mathbf{r})} \frac{\partial \rho_m(\mathbf{r})}{\partial \mathbf{R}_a} d\mathbf{r} \\ &= c_a \int (V_H(\mathbf{r}) + V_m(\mathbf{r})) \frac{\partial g_a(\mathbf{r}, \mathbf{R}_a)}{\partial \mathbf{R}_a} d\mathbf{r}.\end{aligned} \tag{3.9}$$

Since $V_H + V_m$ is only approximately constant within the metal, the integration over $\mathbf{r}$ in the region of the metal will yield very small contributions. The localized Gaussians g_a decay rapidly to zero resulting in a zero-contribution of these terms for regions outside the metal.

3.2.2. Implementation details

The IC method has been implemented in the QM/MM [119, 120] part of the CP2K [121] program package. So far, the QM/MM driver of CP2K provided only an efficient scheme for calculating electrostatic interactions between permanently charged MM atoms and QM charge density. Our implementation is based on the GPW approach [44, 118], which allows to represent the image charge density $\rho_m(\mathbf{r})$ on a regular grid. Within the GPW formalism, this is equivalent to an expansion of the density in an auxiliary basis set of plane waves. The latter allows to solve the Poisson equation in reciprocal space yielding the potential $V_m(\mathbf{G})$, where $\mathbf{G}$ is the reciprocal lattice vector. The potential in real space $V_m(\mathbf{r})$ is obtained by a subsequent inverse Fast Fourier Transform (FFT^{-1}). Periodic boundary conditions (PBC) follow directly from calculating the potential in Fourier space. Note that the QM and MM boxes need to be of the same size, since the KS and IC density have to be represented on commensurate grids. For systems with a small QM region and a very large MM subsystem, this requirement could substantially contribute to the computational cost.

To obtain the density $\rho_m(\mathbf{r})$, the linear set of equations given in Equation 3.5 is solved by employing a Gaussian elimination scheme or a conjugate gradient (CG) procedure. In both cases the integral I_a,

$$I_a = \int (V_H(\mathbf{r}) - V_0) g_a(\mathbf{r}) d\mathbf{r}, \tag{3.10}$$

needs to be computed for each IC center a at every iteration of the SCF cycle. When using a Gaussian elimination scheme, the coefficient matrix $\mathbf{T}$,

$$T_{ab} = \int \int \frac{g_b(\mathbf{r}', \mathbf{R}_b) \cdot g_a(\mathbf{r}, \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r} d\mathbf{r}' = \int v_b(\mathbf{r}) g_a(\mathbf{r}) d\mathbf{r}, \tag{3.11}$$

is calculated following the algorithm illustrated in Figure 3.1. The potential v_b generated by a single Gaussian charge b is obtained after solving the Poisson equation in

a)

Loop over all b atoms

Collocate $g_b(\mathbf{r})$ on the real space grid

Transfer to reciprocal space: $g_b(\mathbf{r}) \xrightarrow{\text{FFT}} g_b(\mathbf{G})$

Solve Poisson equation: $g_b(\mathbf{G}) \longrightarrow v_b(\mathbf{G})$

Transfer back to real space: $v_b(\mathbf{G}) \xrightarrow{\text{FFT}^{-1}} v_b(\mathbf{r})$

Loop over all atoms a with $a >= b$

Calculate integral $\int v_b(\mathbf{r}) g_a(\mathbf{r}) d\mathbf{r}$

End a loop

End b loop

b)

Choose reference atom b_{ref}

Collocate $g_{b_{ref}}(\mathbf{r})$ on the real space grid

Transfer to reciprocal space: $g_{b_{ref}}(\mathbf{r}) \xrightarrow{\text{FFT}} g_{b_{ref}}(\mathbf{G})$

Solve Poisson equation: $g_{b_{ref}}(\mathbf{G}) \longrightarrow v_{b_{ref}}(\mathbf{G})$

Transfer back to real space: $v_{b_{ref}}(\mathbf{G}) \xrightarrow{\text{FFT}^{-1}} v_{b_{ref}}(\mathbf{r})$

Loop over all b atoms

Loop over all atoms a with $a >= b$

shift $g_a(\mathbf{r})$: $\mathbf{R}_{a_{shift}} = \mathbf{R}_a - \mathbf{R}_b + \mathbf{R}_{b_{ref}}$

Calculate integral $\int v_{b_{ref}}(\mathbf{r}) g_{a_{shift}}(\mathbf{r}) d\mathbf{r} = \int v_b(\mathbf{r}) g_a(\mathbf{r}) d\mathbf{r}$

End a loop

End b loop

Figure 3.1.: Pseudocode for calculating matrix $\mathbf{T}$. The inner loop iterates solely over all a larger or equal to atom index b because $\mathbf{T}$ is symmetric. (a) First version: the potential v_b is calculated for each metal atom explicitly. (b) Improved version: the potential v_b is calculated only once for a chosen reference atom.

reciprocal space. After an inverse FFT step, v_b is transferred to real space and the integrals T_{ab} are computed by numerical integration on the real space grid. In the first version of this algorithm published in Ref. [90], the potential v_b is calculated for each metal atom, see Figure 3.1(a). In total, $2N_{\text{IC}}$ FFTs are performed, which can become a bottleneck for the performance for large N_{IC}. Figure 3.1(b) displays an improved version of this algorithm where we utilize that the single Gaussian charges are indeed located at different positions on the grid, but generate the same v_b. The potential v_b is calculated once for a chosen reference atom reducing the number of FFTs to two. In order to obtain the correct integrals, the Gaussian charge at center a has to be accordingly shifted. The number of integrals reduces in both cases to $\frac{N_{\text{IC}}^2}{2} + \frac{N_{\text{IC}}}{2}$, because $\mathbf{T}$ is symmetric. Since matrix $\mathbf{T}$ does not depend on $\rho(\mathbf{r})$, it needs to be

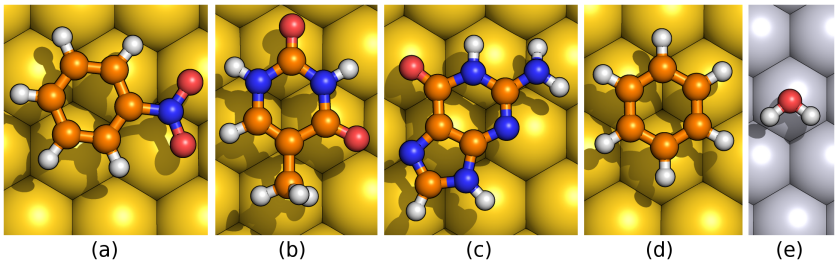

Figure 3.2.: DFT optimized structures of the five investigated molecules adsorbed on a metallic surface: (a) nitrobenzene, (b) thymine, (c) guanine, (d) benzene on Au(111) and (e) water on Pt(111). Color code: *orange*: C, *white*: H, *blue*: N, *red*: O, *yellow*: Au, *gray*: Pt.

calculated only once at the very beginning of the SCF cycle. However, for large N_{IC}, computing the integrals T_{ab} on the full grid can become computationally expensive.

The iterative CG scheme can be used as alternative algorithm to solve the set of linear equations. The expansion coefficients c_a are initially guessed and ρ_m and V_m are calculated. The CG procedure is then applied to update the coefficients iteratively until the electrostatic energy is minimized. At each iteration of the CG, only two FFTs and the calculation of N_{IC} numerical integrals are required, which can, in principle, be computationally more efficient than the exact calculation of matrix $\mathbf{T}$. However, the CG optimization requires several iterations that need to be performed at every SCF step. To ensure a fast convergence of the CG, a proper preconditioning of the gradient is applied. The optimal choice for the preconditioner is a good approximation of the matrix $\mathbf{T}$. In practice, we always compute $\mathbf{T}$ in the first step. Geometry optimizations and MD simulations can exploit the fact that the same matrix $\mathbf{T}$ can be used as CG preconditioner for several configurations before being updated. Therefore, the Gaussian elimination scheme is always the method of choice for single point calculations. The preconditioned-CG is used for geometry optimizations and MD simulations.

Our implementation is designed to enable simulations of large systems and takes advantage of the massive parallelization procedures already present in the CP2K program package.

3.3. Computational details

3.3.1. Static simulations

System setup

The IC-QM/MM approach has been tested on five different adsorbate-metal systems involving physi- and chemisorbed molecules. These are benzene, nitrobenzene,

thymine and guanine adsorbed on Au(111) as well as water adsorbed on Pt(111). The DFT optimized structures are displayed in Figure 3.2. In the case of H_2O/Pt, a water dimer and a cluster of 12 molecules have also been considered. The metallic substrates are modeled by a four-layer slab and PBC are applied in all three dimensions. In order to properly accommodate the adsorbates on the surface, avoiding spurious interactions with the periodic images, $p(12\times12)$ replica of the unit cell are used in the lateral dimensions. To decouple the periodic images in the dimension perpendicular to the surface, a vacuum of 25 Å is added along the [111] direction. The slab structure is asymmetric, meaning that the molecule is adsorbed only on one side of the slab, while the atoms of the lowest metal layer are kept fixed in the bulk positions.

QM part of the calculation

Full DFT calculations are performed as reference and compared to QM/MM results obtained with and without IC correction. All calculations are carried out with the CP2K program package [121]. For QM calculations, double-ζ valence plus polarization (DZVP) basis sets of the Molopt-type [122] are used to represent the valence electrons. The interactions between valence and core electrons are described by norm-conserving Goedecker, Teter and Hutter pseudopotentials [123–125]. The Perdew-Burke-Ernzerhof (PBE) functional [20] models the exchange and correlation potential. Van der Waals interactions are accounted for by employing Grimme's D3 dispersion correction [34]. The energy cutoff for the auxiliary plane wave expansion of the density is set to 400 Ry.

MM part of the calculation

At the MM level of theory, Au and Pt are described through embedded atom model (EAM) potentials [126] obtained from the LAMMPS [127] data base. Lattice constants were optimized at the respective level of theory. For Au, lattice parameter of 4.08 Å (EAM) and 4.20 Å (DFT) are used, whereas the experimental spacing of 3.92 Å is employed for both, EAM and DFT, in case of Pt. Using the same lattice constants for EAM and DFT is possible for Pt, since the DFT spacing deviates by less than 2% from the experimental (EAM) one.

The MM-based interactions between metal and molecule are modeled with empirical potentials reproducing dispersion, Pauli repulsion and, for H_2O/Pt, specific chemisorption terms. Induction effects are accounted for by image charges. A modified Born-Mayer potential,

$$V(r_{ij}) = A_{ij}\exp(-r_{ij}/\kappa) - C_{ij}r_{ij}^{-6}\cdot f_{\mathrm{dmp}}(r_{ij}), \tag{3.12}$$

is employed for the interactions between Au and adsorbed molecules. The parameters were taken from Ref. [128]. The attractive term was modified with a damping

function f_{dmp} as used in Ref. [33] in order to reproduce the full DFT molecule-surface separation. This potential is not expected to reproduce accurate structures, but is only instrumental to estimate the effects of the IC method on a standard QM/MM setup.

Interactions between water molecules and Pt are computed from the Siepmann-Sprik potential [104], which has been implemented in the MM module of CP2K. The chemisorption potential by Siepmann and Sprik is a sum of two- and three body terms. It was parametrized to correctly model the ground state geometry and potential curves determined by *ab initio* cluster calculations. The two-body part of the chemisorption potential defines the pair interactions between Pt and O atoms,

$$v_2(r_{ji}) = Ar_{ji}^{-\alpha} - Cr_{ji}^{-6} - Df_2(r_{ji})r_{ji}^{-3}\Phi_{ji}(\phi_{ji}), \tag{3.13}$$

where r_{ji} is the distance between the oxygen of water molecule i and Pt atom j. The first two terms describe repulsion and attractive dispersive interactions, respectively. The third term depends additionally on the angle ϕ_{ji} between the H_2O dipole vector of molecule i and vector $\mathbf{r}_{ji}$. The purpose of this term is to capture chemical bonding. Since this is only required at the interface, we modulate with a short-range cut-off function,

$$f_2(r_{ji}) = \exp\left[\frac{B}{r_{ji} - r_{cut}}\right] \tag{3.14}$$

which is set to zero for $r_{ji} \geq r_{cut}$. The function $\Phi_{ji}(\phi_{ji})$ directs the hydrogens away from the surface and is given by

$$\Phi(\phi) = \exp\left[-8\left(\frac{\cos\phi - 1}{4}\right)^4\right]. \tag{3.15}$$

The three-body function defines the position of the water molecules on the surface and pushes the water on top of the metal atoms,

$$v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = Ef_2(r_{ji})r_{ji}^{-\beta}\exp\left[F\cos^2\left(\frac{\theta_{ijk}}{2}\right)\right]. \tag{3.16}$$

θ_{ijk} is the angle between $\mathbf{r}_{ji}$ and the vector $\mathbf{r}_{jk}$ between Pt atom j and its neighbors k. The v_3 term is entirely repulsive and is sensitive to different surface topologies. The parameters $A, B, C, D, E, F, r_{cut}, \alpha$ and β are given in Ref. [104]. Further details on the implementation and the expressions for the corresponding forces can be found in Appendix A.

Analysis of properties

Different structural and electronic properties have been analyzed for the optimized structures. In particular, interaction energies E_{int} are calculated according to

$$E_{\mathrm{int}} = \frac{E_{\mathrm{slab+mol}} - (E_{\mathrm{slab}} + E_{\mathrm{mol}})}{n}, \tag{3.17}$$

where $E_{\mathrm{slab+mol}}$ denotes the total energy of the optimized adsorbate-metal system and n the number of adsorbed molecules. E_{slab} and E_{mol} are the energies of the slab and molecules in the geometry of the complex, respectively. Adsorption energies are instead computed from the energies of the optimized individual components as

$$E_{\mathrm{ads}} = \frac{E_{\mathrm{slab+mol}} - (E_{\mathrm{slab}}^{\mathrm{opt}} + nE_{\mathrm{mol}}^{\mathrm{opt}})}{n}. \tag{3.18}$$

The interaction and adsorption energies obtained from full DFT calculations are affected by the basis set superposition error and have been corrected by the counterpoise formalism [129, 130].

Electron density differences $\Delta\rho_{\mathrm{elec}}$ are used to investigate changes in the molecular electronic distribution induced by the metal. $\Delta\rho_{\mathrm{elec}}$ is obtained as grid-based difference of $\rho_{\mathrm{elec}}^{\mathrm{slab+mol}}$ and the electronic densities of the individual components, taken with atomic coordinates as in the complex,

$$\Delta\rho_{\mathrm{elec}} = \rho_{\mathrm{elec}}^{\mathrm{slab+mol}} - \rho_{\mathrm{elec}}^{\mathrm{slab}} - \rho_{\mathrm{elec}}^{\mathrm{mol}}. \tag{3.19}$$

For purpose of representation, we introduce the one-dimensional electron density difference integrated in the planes parallel to the surface,

$$\Delta\rho_{\mathrm{elec}}^{\mathrm{1D}} = \int \Delta\rho_{\mathrm{elec}}\, dxdy. \tag{3.20}$$

Furthermore, molecular dipole moments are computed from the positions of the atomic cores and the centers of maximally localized Wannier orbitals (MLWO) [131, 132].

3.3.2. MD simulation

IC-QM/MM has also been applied to perform an MD simulation of the liquid water/Pt(111) interface. The supercell of the substrate is modeled laterally with a $p(6\times8)$ repetition of the unit cell and vertically by four metal layers. The Pt(111) surface is surmounted by a 15 Å thick liquid water film, consisting of 151 molecules. Above the water system, 20 Å of vacuum space are introduced to decouple the periodic images along the vertical axis. The QM/MM setup is identical to the setup of the static water/Pt(111) calculations. The simulation is performed within the microcanonical (NVE) ensemble at a temperature of 330 K and a time step of 1 fs is used. The system has been equilibrated for 25 ps employing velocity rescaling for the first 18 ps. The last 55 ps of the simulation have been used for analysis of structural features. The measured drift of the total energy is in the range of 10^{-5} eV/ps/atom during the production run. MLWO were calculated every fifth time step for the last 40 ps of the simulation and used for calculation of the average molecular dipole moments. For analysis of the electronic rearrangement, $\Delta\rho_{\mathrm{elec}}$ has been calculated and averaged for ten equally spaced snapshots.

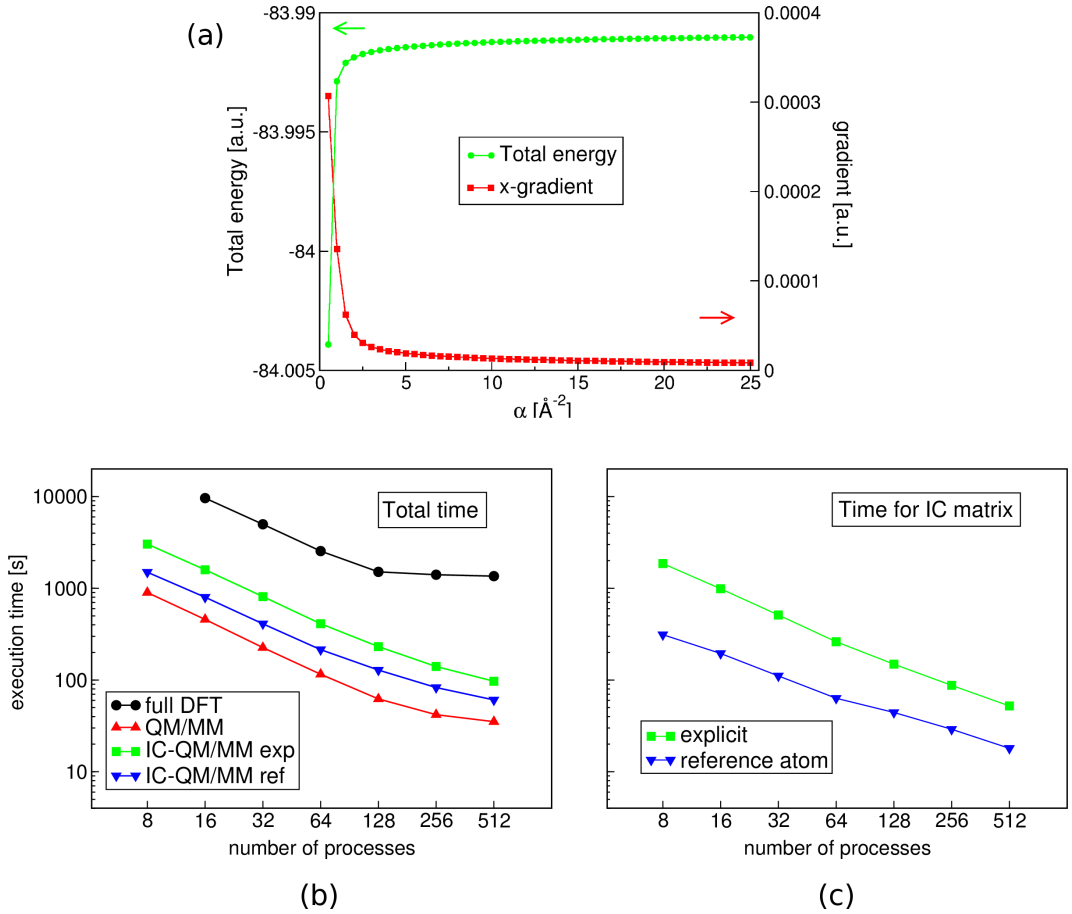

Figure 3.3.: Technical parameters. (a) Dependence of total energy and x-gradients (of an arbitrary metal atom) on parameter α. The x-gradients refer to the IC contribution, see Equation 3.9. (b) Execution time of a single point calculation of thymine on Au(111) performed on a Cray XC30 system. Note that a double logarithmic scale is employed. (c) Time for calculating IC matrix $\mathbf{T}$. The labels *explicit* and *reference atom* refer to the two versions of the algorithm displayed in Figure 3.1.

3.4. Tests and applications

3.4.1. Gaussian width α and performance

The only adjustable variable in the IC-QM/MM scheme is the width α of the Gaussian charge distribution ρ_m, see Equation 3.1. To assess the influence of α on the electrostatic interaction energy, test calculations using the thymine-gold system have been performed. The dependence of total energy and gradients on the Gaussian width is reported in Figure 3.3(a). For values of α larger than 3.0 Å^{-2}, neither the total energy nor the image charge contribution to the gradients show a significant dependency on α. However, drastic changes are found for smaller values. Small values of α correspond to very broad Gaussians, which results in a uniform charge distribution across the metal, i.e. the Gaussians are no longer localized at the atomic centers. This leads

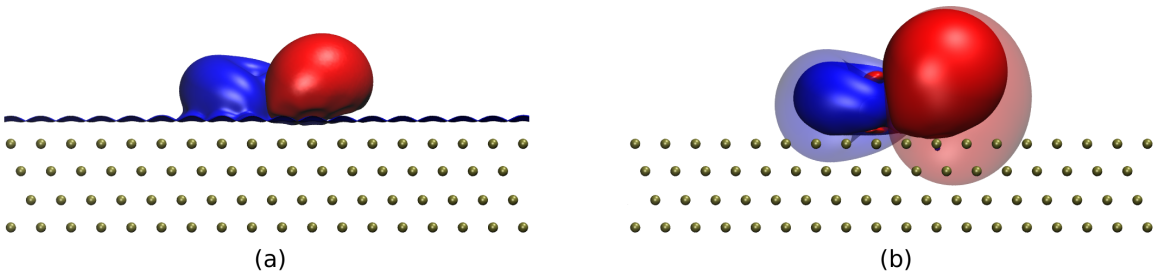

Figure 3.4.: Nitrobenzene on Au(111). Isosurface of the electrostatic potential obtained with (a) full DFT, (b) IC-QM/MM (solid) and (b) QM/MM (transparent). For IC-QM/MM, the electrostatic potential corresponds to $V_m + V_H$ as defined in Equation 3.2. Red: vacuum value + 0.004 a.u. Blue: vacuum value -0.004 a.u.

obviously to non-physical results arising from technical artifacts. Using extremely large values for α is also not recommendable because very steep Gaussians are not accurately described by a mesh representation over finite grids. For the calculations discussed in the following, a value of $3.0\,\text{Å}^{-2}$ has been chosen.

In order to evaluate the performance of our IC-QM/MM implementation, the average time needed for a single point calculation for thymine on Au(111) has been estimated, see Figure 3.3(b). Using the first version of the algorithm shown in Figure 3.1(a), the additional computational cost introduced by the IC correction is mainly determined by the evaluation of matrix $\mathbf{T}$. The test system consists of 576 metal atoms resulting in a matrix of size 576×576. Consequently, FFT has to be performed 1152 times and 166 176 integrals have to be calculated. Employing the improved version of the algorithm, the single Gaussian potential is only calculated for a chosen reference atom reducing the number of FFTs to two. This results in a speed-up of factor 5-6 for calculating the matrix $\mathbf{T}$, see Figure 3.3(c). For the total execution time, a factor of two is gained. The parallel efficiency of the IC-QM/MM scheme is shown in Figure 3.3(b). For this system, the scaling is good up to 512 cores, while beyond this limit no further speed-up is possible due to load-balancing effects. The comparison of timings reported in Figure 3.3(b) also shows that by introducing the IC approximation with the improved algorithm, the QM/MM calculations become approximately two times slower. However, IC-QM/MM is still up to 22 times faster than a full DFT calculation.

3.4.2. Electrostatic potential and induced charges

Isosurfaces of the electrostatic potential as obtained by full DFT, IC-QM/MM and QM/MM are displayed for nitrobenzene on Au(111) in Figure 3.4. The electrostatic potential V_H generated by the adsorbed molecule in the QM/MM calculation extends inside the metallic substrate. By applying the IC correction, the screening of the metal is imposed. As mentioned before, $V_H + V_m$ is not required to be exactly zero,

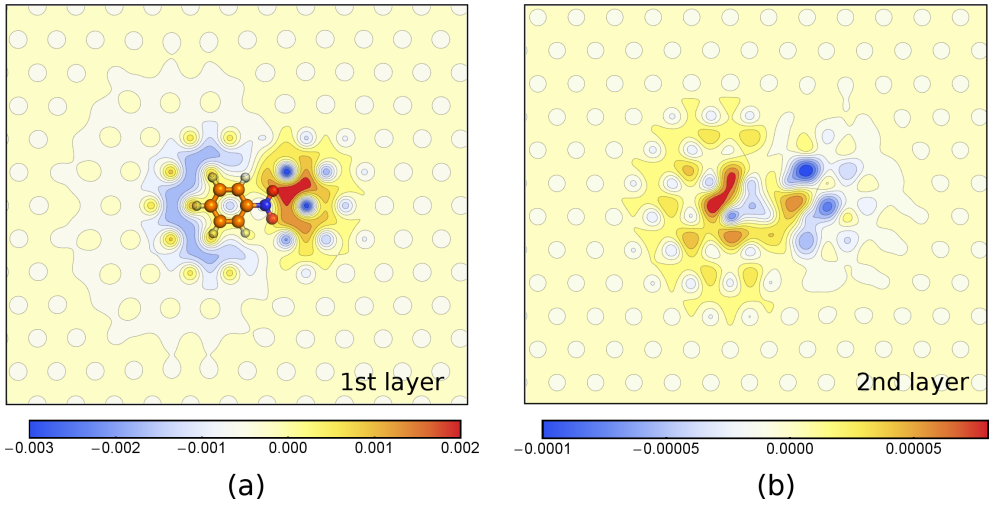

Figure 3.5.: Nitrobenzene on Au(111). Contour plot of the electrostatic potential ($V_m + V_H$) [a.u.] in the first (a) and second (b) layer of the metal slab as obtained with IC-QM/MM.

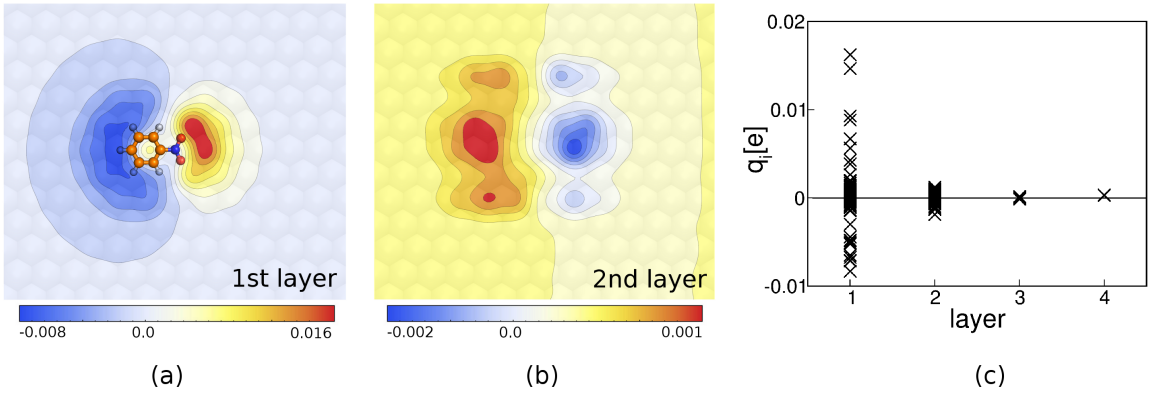

Figure 3.6.: Image charges induced in a 4-layer Au(111) slab by a nitrobenzene molecule. Charge distribution induced in the first (a) and second (b) metal layer. Magnitude q_i of the induced charges in layer 1-4 (c).

but only its weighted integral. However, the sum of $V_H + V_m$ is basically zero at every grid point $\mathbf{r}$ inside the slab, as evident from the isosurfaces of the electrostatic potential shown in Figure 3.4(b). The contour plots of the potential provide more details, see Figure 3.5. At some grid-points, very small non-zero values can occur. Comparing IC-QM/MM to full DFT, qualitatively similar potentials are obtained. However, the potential in the immediate vicinity (i.e. within the Van der Waals radii) of the metallic centers of the first layer is not correctly reproduced by IC-QM/MM. The reason is that IC-QM/MM does not take into account the explicit electronic structure of the substrate neglecting the spatial expansion of the electron shell.

Despite the fact that image charges are introduced to capture a macroscopic effect, investigation of induced charges at an atomistic level reveals some interesting features. The IC distribution induced by nitrobenzene on Au(111) is displayed in the contour

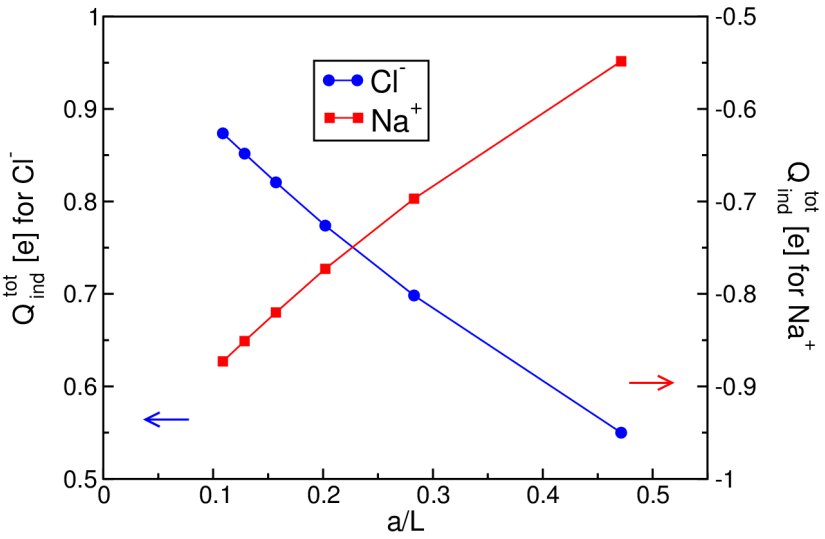

Figure 3.7.: Dependence of the total induced charge $Q_{\text{ind}}^{\text{tot}}$ on the size of a finite one-layer Au(111) slab. The inverse length L is given in units of the lattice constant a. The charged ions Na^+ and Cl^- have been placed 4.0 Å above the metal slab. The computational setup corresponds to the one described in Section 3.3 except that PBC are not applied.

plots of Figure 3.6(a) and (b). The induction effect occurs mainly in the first layer, while its magnitude decreases rapidly with the distance from the surface. Compared to the first layer, charges induced in the second layer are an order of magnitude smaller. In the third and fourth layer, the induction effect essentially vanishes, see Figure 3.6(c). It is further observed that, as expected, the electronegative *nitro* group induces positive charges and the electropositive benzene ring induces negative image charges in the first layer. The inverse distribution is found in the second layer.

The total sum of the image charges for the given test system is in the order of 10^{-2}. This sum is expected to converge to zero when reaching the macroscopic continuum limit [104, 107], i.e. in the limit of an infinite metal slab. The finite sum of the image charges does not correspond to a physical charge since the image charges are merely a computational tool to impose the correct behavior of the electrostatic potential. This is demonstrated by placing Na^+ and Cl^- ions above a finite metal slab, where a total induced charge of -1 is expected for Na^+ and $+1$ for Cl^-. As shown in Figure 3.7, the total induced charges approach these values with increasing size of the slab.

3.4.3. Aromatic adsorbates on Au(111)

Geometries and energetics

For nitrobenzene, thymine and guanine adsorbed on Au(111), full DFT structure optimizations yield basically a planar geometry of the adsorbate with an adsorbate-surface separation $d_{\text{ad-metal}}$ of about 3.1 Å. An almost planar structure has been ob-

Table 3.1.: Interaction E_{int} (kJ/mol) and adsorption energies E_{ads} (kJ/mol) of thymine and guanine on Au(111). Energies are given for the structure optimized with the respective method.

	thymine		guanine	
	E_{int}	E_{ads}	E_{int}	E_{ads}
QM/MM	−59.4	−58.3	−67.0	−65.5
IC-QM/MM	−63.3	−62.3	−73.1	−71.4
$\Delta_{IC-QM/MM}$	−3.9	−4.0	−6.1	−5.9

tained for benzene/Au(111) with benzene adsorbed at an fcc threefold hollow site and $d_{ad-metal}$=3.0 Å, which is in agreement with the results reported in Ref. [133]. High symmetry adsorption sites are not observed for nitrobenzene, thymine and guanine. The optimization of the molecular structures by QM/MM also results in planar geometries and the metal-molecule separations are in agreement with the DFT ones. The QM/MM structures are almost unchanged upon the introduction of image charges, which is attributed to the relatively large molecule-substrate separations. The reason is that the induction effect is relatively weak and the related force contributions are too small to noticeably change the structure.

The interaction and adsorption energies are computed for thymine and guanine and reported in Table 3.1. The contributions to interaction and adsorption energies arising from polarization lies in the range of −4 to −6 kJ/mol, which coincides with the values given in Ref. [104]. Comparing thymine and guanine, it appears that the polarization energy is larger for the latter. This is attributed to the larger dipole moment (see below) of guanine causing a stronger induction effect. As expected, the IC correction strengthens the interaction between the adsorbate and metal. Nonetheless, energies obtained with full DFT deviate up to −40 kJ/mol from the QM/MM results. Having chosen existing MM interaction pair potentials, not parametrized for the specific model described in this work, we cannot expect a better agreement with the DFT values. Our goal is the assessment of the IC approach. The development of better force fields for these test cases is beyond the scope here.

Even better understanding of the energetic aspects of the interaction with the induced charges is provided by the variation of the molecule-surface separation. In the plots displayed in Figure 3.8, z is the height of the molecule above the surface, and z_{eq} is the height corresponding to the optimized geometry. Panel (a) of Figure 3.8 shows the variation of the total QM energy E_{QM} and the image energy E_{IC} of thymine adsorbed on Au(111). E_{IC} is the energy of the molecule derived from the interaction between the molecular charge density and the IC distribution (second term of Equation 3.6). This energy term decreases (in absolute terms) monotonically by increasing the distance. The modulation of E_{QM} is due to the presence of the image charges,

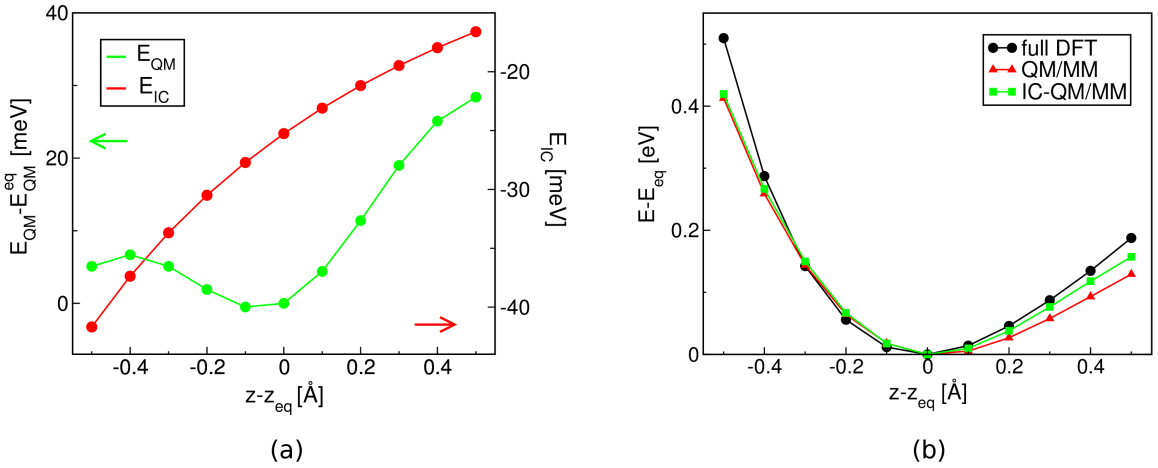

Figure 3.8.: Potential curves. (a) Image energy E_{IC} and difference in total QM energy E_{QM} vs. thymine-Au(111) separation for IC-QM/MM. (b) Difference in total energy E vs. thymine-Au(111) separation. E_{eq} and $E_{\mathrm{QM}}^{\mathrm{eq}}$ are the energies at equilibrium distance z_{eq}. The potential curves have been calculated for the structure optimized with the respective method.

which induce a self-consistent perturbation of the electronic structure of the molecule. In plain QM/MM calculations, E_{QM} of the molecule is not affected by the separation from the metal. If the self-consistent treatment would be omitted, E_{QM} would display the same behavior as E_{IC} and a minimum at z_{eq} could not be observed.

The variation of the total energy E vs. the distance from the substrate are reported in panel (b) of Figure 3.8. For separations larger than the equilibrium distance, the image charges significantly increase the steepness of the QM/MM curve and approximate the full DFT curve better. Including polarization accounts in this case for roughly half of the difference between full DFT and QM/MM. Considering distances between z_{eq} and $z > -0.4\,\text{Å}$, all three curves coincide. For values smaller than $z = -0.4\,\text{Å}$, the full DFT curve is steeper indicating repulsive interactions that are not captured by our relatively simple polarization model. Since image charges are modeling a long-range effect, it is reasonable that improvements are obtained for $z > z_{\mathrm{eq}}$, but only very limited for $z < z_{\mathrm{eq}}$.

Electronic structure

In the self-consistent procedure, the induced charge density ρ_m and the molecular density ρ mutually modify each other. In order to study the electronic structure rearrangements of the adsorbate, we investigated the difference in electron density $\Delta\rho_{\mathrm{elec}}$ upon adsorption of benzene on Au(111). This system is an ideal test candidate since, due to its planar and highly symmetric structure, polarization is only occurring vertically to the molecule plane. Figure 3.9(a) shows $\Delta\rho_{\mathrm{elec}}$ for benzene as obtained with IC-QM/MM. Accumulation of electron density (red) is observed below

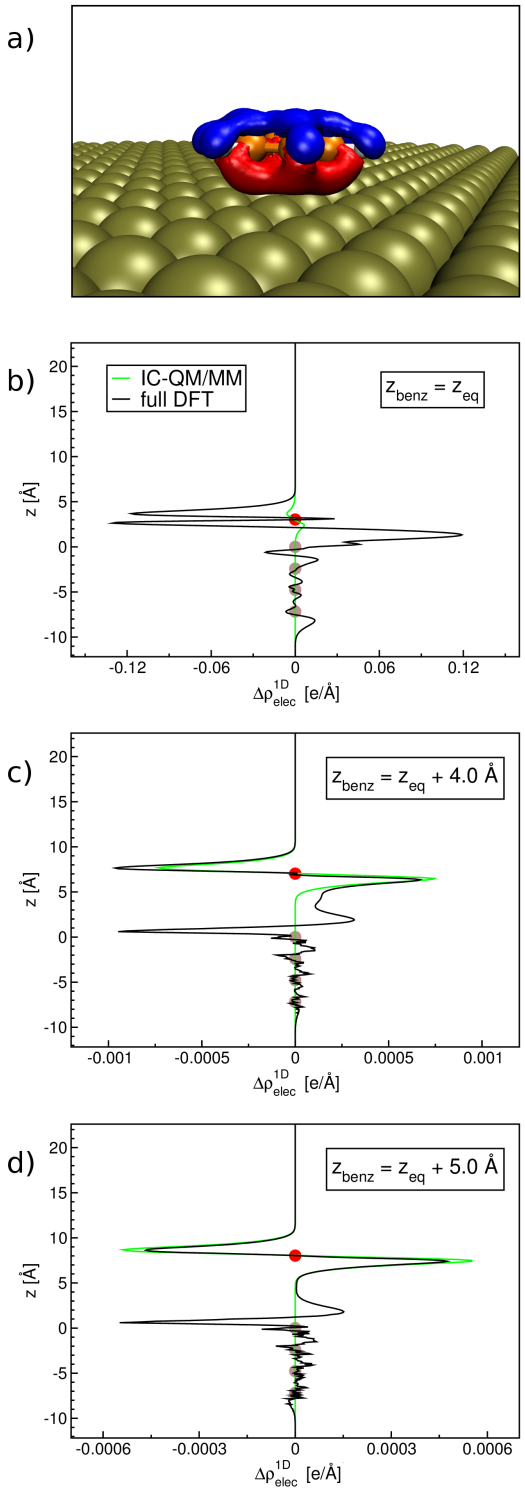

Figure 3.9.: Difference in electron density $\Delta\rho_{\mathrm{elec}}$ upon adsorption of benzene on Au(111). (a) Three-dimensional representation of $\Delta\rho_{\mathrm{elec}}$ obtained with IC-QM/MM. Red regions correspond to increased and blue regions to decreased electronic density. (b-d) Plane-integrated density difference $\Delta\rho^{\mathrm{1D}}_{\mathrm{elec}}$. Brown spheres indicate the position of the metal layers and the red sphere the position z_{benz} of benzene. z_{eq} indicates the equilibrium separation obtained by full DFT.

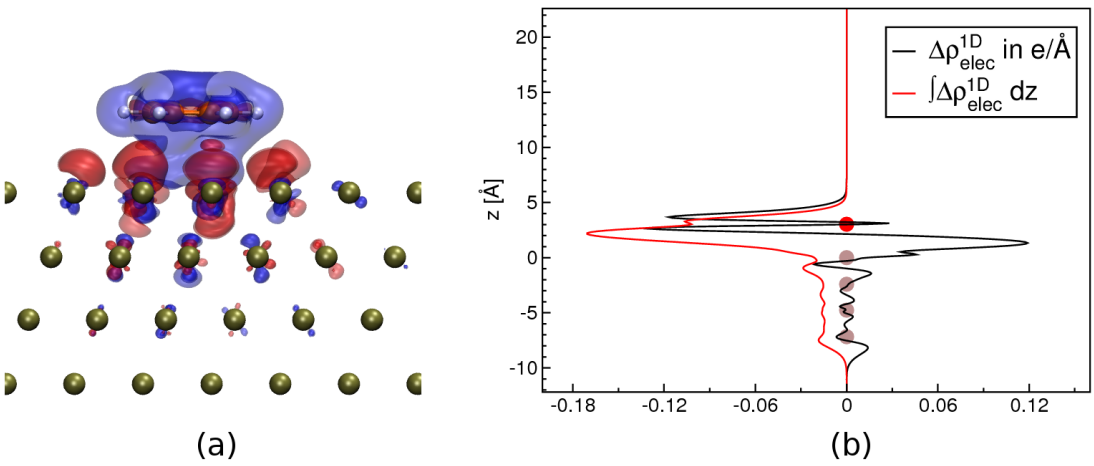

Figure 3.10.: Difference in electronic density $\Delta\rho_{\text{elec}}$ upon adsorption of benzene on Au(111) as obtained by full DFT calculations. (a) Three-dimensional plot. Red regions correspond to increased and blue regions to decreased electronic density. (b) Plane-integrated electron density difference $\Delta\rho^{\text{1D}}_{\text{elec}}$. Brown spheres indicate the position of the metal layers and the red sphere the position of the benzene molecule. The cumulative or "running" integral of $\Delta\rho^{\text{1D}}_{\text{elec}}$ (starting integration from the vacuum) corresponds to the number of electrons lost (negative) or gained (positive).

and depletion (blue) above the ring plane. The vertical rearrangement of the electron density points at a polarization of the π-electron system towards the surface. This is also clearly appearing in the plots of the plane-integrated density difference $\Delta\rho^{\text{1D}}_{\text{elec}}$, see Figure 3.9(b-d). Negative values of $\Delta\rho^{\text{1D}}_{\text{elec}}$ correspond to depletion and positive values to accumulation of electrons. Figure 3.9(b) presents $\Delta\rho^{\text{1D}}_{\text{elec}}$ for benzene located at its equilibrium position z_{eq} as obtained by full DFT. Close to the position of the benzene (red sphere), IC-QM/MM shows the expected negative peak above and a positive peak below the ring plane. However, $\Delta\rho^{\text{1D}}_{\text{elec}}$ obtained by full DFT is by factor of 10 larger than predicted by IC-QM/MM and the shape of the peaks is more sophisticated. As obvious from the one- and three-dimensional representations of $\Delta\rho_{\text{elec}}$ (see also Figure 3.10(a)), the density between benzene and substrate is not separated. Bader charge analysis [134, 135] implies a charge transfer of 0.14 electrons from benzene towards the surface. The cumulative integral of $\Delta\rho_{\text{elec}}$, displayed in Figure 3.10(b), shows the loss of electrons vertically to the surface plane and confirms the value obtained by the Bader charges. From that we can conclude that the molecule is not solely physisorbed on the surface, but that considerable electronic interactions between benzene and Au(111) occur. This type of interactions at short distances cannot be reproduced by polarization models like the IC approach alone.

By rigidly displacing benzene to larger distances from the metal, the overlap between the surface and molecular states is reduced. At large distances, the only perturbation remaining must be attributed to polarization. As demonstrated by the plots in

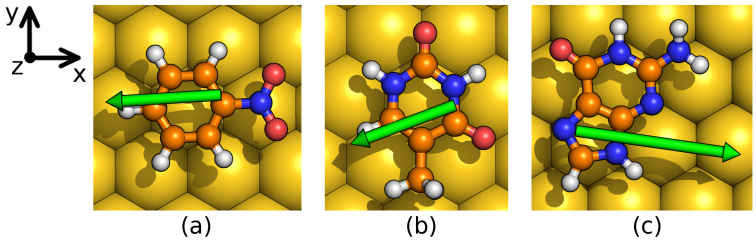

Figure 3.11.: Dipole moments (green arrows) of the adsorbates with atomic coordinates optimized by IC-QM/MM. (a) Nitrobenzene, (b) thymine and (c) guanine on Au(111).

Table 3.2.: Dipole moments μ (Debye) for nitrobenzene, thymine and guanine adsorbed on Au(111). μ_x, μ_y, μ_z are the x-, y-, and z-component of the dipole vector respectively. IC-QM/MM structures have been used for all adsorbates.

method	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	$\Delta\lvert\mu\rvert$
			nitrobenzene		
QM/MM	-4.59	-0.29	-0.25	4.61	0.15
IC-QM/MM	-4.75	-0.30	-0.25	4.76	
			thymine		
QM/MM	-4.02	-1.50	0.06	4.29	0.17
IC-QM/MM	-4.19	-1.55	0.06	4.46	
			guanine		
QM/MM	6.53	-0.97	0.35	6.61	0.23
IC-QM/MM	6.75	-1.03	0.43	6.84	

panels (c) and (d) of Figure 3.9, the agreement between DFT and IC-QM/MM in the region of the adsorbate is also quantitatively very good in this case. By construction, the electronic structure of the metal cannot be reproduced within the IC model.

To further study electronic effects due to polarization, molecular dipoles have been calculated from the position of the Wannier centers as obtained for the structures optimized with the IC-QM/MM method. Without the IC contributions, the electronic structure of the molecule is not affected by the presence of the metal. Therefore, the electronic properties are expected to be those of the molecule in gas phase, apart from small modifications induced by the slight geometrical rearrangement of the physisorbed system. The polarization of the electronic structure through the IC potential is, instead, well recognized in the increased dipole moments of all three tested molecules, see Table 3.2. The single components of the dipole vectors μ_x, μ_y and μ_z refer to the molecule and axis orientation displayed in Figure 3.11. Since the adsorbed molecules are almost planar, the dominant components of the dipoles are those in the xy plane. The largest changes of the dipole due to IC polarization is also observed within this plane. This corresponds to an enhancement of the charge separation within

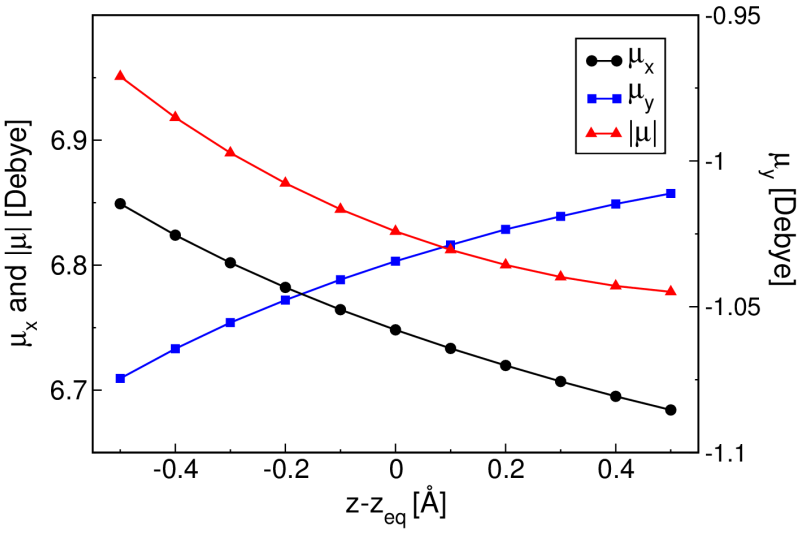

Figure 3.12.: Variation of the dipole moment μ in Debye for guanine adsorbed on Au(111) dependent on the molecule-surface separation. μ_x and μ_y are the x- and y-component of the dipole vector respectively. z_{eq} is the equilibrium position of the molecule as obtained from the IC-QMMM model.

the molecule between the more electronegative and the more electropositive groups. Only in the case of guanine, which already has a positive μ_z component, we observe a non-negligible increment of the dipole also along the z-axis.

QM/MM calculations neglecting induction cannot capture the effect that the dipole moment is sensitive to the molecule-surface separation. Displacing the molecule along the z-axis, the stronger or weaker interaction with the image charges causes a modulation of the molecular dipole. Figure 3.12 exemplary displays distance-dependent alterations of the lateral component of the dipole moment as well as its magnitude for guanine on Au(111). We observe that the enhancement induced by the image charges decays as the inverse of the distance from the metal, which is indeed the asymptotic long-range behavior of the IC potential. For large separations, the dipole moment converges to the value for an isolated guanine molecule or in other words, to the QM/MM value without image charges. This is demonstrated in Figure 3.13 where the lateral and vertical components of the dipole moment are displayed for large displacements of up to 2Å from the equilibrium position.

3.4.4. Water on Pt(111)

Static calculations of water clusters

Starting with the geometry of a single water molecule on Pt(111), the molecule-surface separation predicted by full DFT is with 2.38 Å considerably smaller than for aromatic adsorbates on Au(111). Hence, stronger induction effects are expected

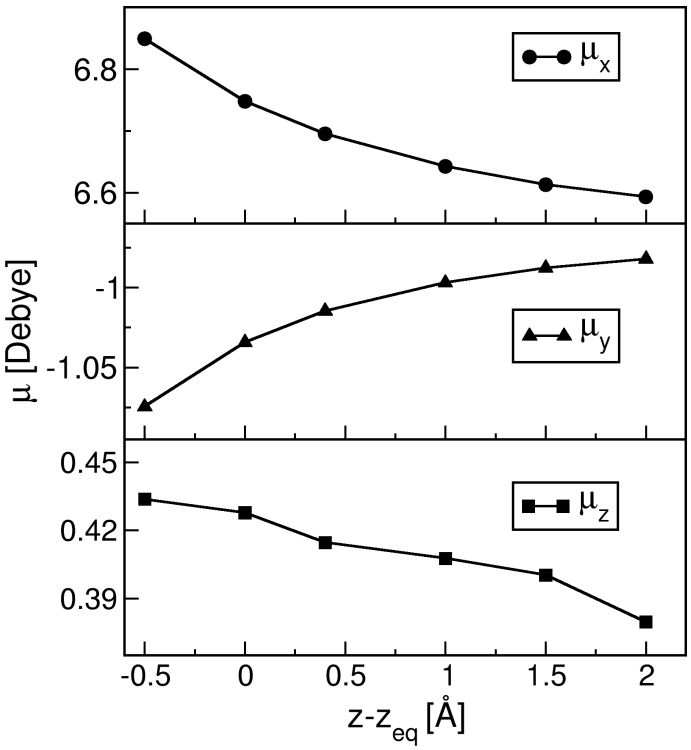

Figure 3.13.: Variation of the dipole moment μ in Debye for guanine adsorbed on Au(111) for large molecule-surface separations. μ_x, μ_y and μ_z are the x-, y- and z-component of the dipole vector respectively. z_{eq} is the equilibrium position of the molecule as obtained from the IC-QM/MM model.

potentially causing structural changes. While the distance is essentially the same with all three models, the induced charges play an important role in determining the tilt angle. The latter is defined as the angle between surface normal and dipole vector of the H_2O molecule. According to full DFT results, the H_2O monomer binds preferentially at atop sites and lies nearly flat on the surface with a tilt angle of roughly 75°. Similar results were reported in Ref. [136]. Atop binding is also observed with the QM/MM potential describing the interactions between water and platinum. The empirical potential was designed to rotate the molecule such that the hydrogens point up. That means, a small angle between surface normal and dipole moment is energetically favored. The angle predicted by full DFT is roughly 60° larger indicating an attractive interaction between hydrogens and platinum. IC-QM/MM yields an angle that is approximately 10° closer to full DFT. Despite the fact, that the difference to full DFT is still large, image charges account at least partly for the rotation of the tilt angle. This demonstrates that polarization effects are important for modeling the interactions between hydrogen and platinum. Structural differences become much more pronounced when several water molecules are adsorbed on Pt(111). The changes in geometry compared to QM/MM without image charges are reported in Figure 3.14. The most remarkable effect of the image charges is that water is not only adsorbed

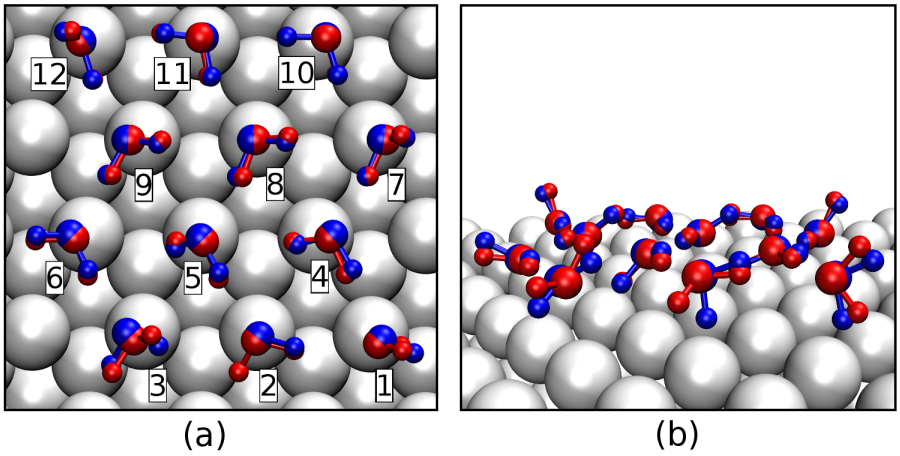

Figure 3.14.: Top view (a) and perspective (b) view of the optimized structures of the $12H_2O$ water adsorbed on Pt(111). *Red:* structure optimization using QM/MM without IC. *Blue:* structure optimization using IC-QM/MM.

via metal-oxygen (*M-O*), but also via metal-hydrogen (*M-HO*) bonds. "H-down" water molecules are also predicted by full DFT structure optimization. Experimental studies [137–139] using X-ray absorption, X-ray emission and X-ray photoelectron spectroscopy found, indeed, that water is adsorbed (intact) with both *Pt-O* and *Pt-HO* bonding species on Pt(111). More recent studies of water monolayers on Pt(111) employing scanning tunneling [140] and vibrational spectroscopy [141] found a so-called $\sqrt{39}$ structure of the monolayer. Nonbonded hydrogens in this structure show also the preference to point towards the surface rather than away.

The interaction and adsorption energies of the water clusters are reported in Table 3.3. The contribution of the image charges to the interaction energies is roughly $-3\,\mathrm{kJ/mol}$ for $1H_2O$ and $2H_2O$, which agrees with the values reported in Ref. [104]. For $12H_2O$, the contribution of the image charges is two times larger indicating that the overall polarization effect is increased with increasing number of adsorbates. Generally, the IC-QM/MM energies are close to the energies predicted by full DFT. Especially, the adsorption energies obtained with IC-QM/MM show good agreement with the full DFT results. The deviation is less than $4.0\,\mathrm{kJ/mol}$ for all water systems. Comparing interaction and adsorption energies, it is to note that the latter contain also intermolecular hydrogen bonding E_{H-bond} resulting in more negative values for the adsorption energies of the clusters.

The increase in dipole moment is almost marginal for a single molecule upon introduction of polarization. However, the changes in electronic structure are much more pronounced for a whole water cluster. Table 3.4 summarizes the results for $12H_2O$. The dipole moments obtained with QM/MM and IC-QM/MM have been calculated for the respective structures presented in Figure 3.14. The numbering of the molecules corresponds to the one shown in Figure 3.14(a). The dipole moment of

Table 3.3.: Interaction E_{int} (kJ/mol) and adsorption energies E_{ads} (kJ/mol) of water clusters on Pt(111). E_{ads} includes intermolecular hydrogen bonding E_{H-bond} (kJ/mol per molecule). Energies are given for the structure optimized with the respective method.

	$1\,H_2O$		$2H_2O$			$12H_2O$		
	E_{int}	E_{ads}	E_{int}	E_{ads}	E_{H-bond}	E_{int}	E_{ads}	E_{H-bond}
QM/MM	−41.6	−37.3	−40.9	−49.2	−10.6	−36.4	−61.9	−26.0
IC-QM/MM	−44.2	−43.6	−43.7	−52.9	−10.5	−42.8	−66.6	−24.4
full DFT	−44.9	−43.5	−50.6	−56.8	−7.0	−44.2	−63.0	−19.7

a single water molecule in the cluster can be up to 0.7 Debye larger than in gas phase (1.85 Debye [142]) due to the formation of hydrogen bonds (H-bonds) with neighboring water molecules. Significantly larger dipole moments are already obtained by the QM/MM description omitting polarization. Several theoretical [143, 144] studies have already exhaustively discussed this phenomenon. For reference, DFT simulations with the PBE functional predict average dipole moments of bulk water between 3.10 to 3.27 Debye [143,145,146]. In addition to that, the effect of the IC polarization induces a further enhancement of the molecular dipoles up to 0.3 Debye. In particular, molecule 3 undergoes the largest dipole-change. Namely, by introducing the IC polarization, molecule 3 changes its orientation and moves towards molecule 6 (see Figure 3.14(a)), strengthening the corresponding H-bond. In general, the changes in dipole moments are due to both, the structural changes and the difference in the electrostatic potential given by the image charges.

MD simulation of a water film

A typical snapshot of the water film on Pt(111) and the first adsorbed water layer (adlayer) are presented in Figure 3.15(a) and (b). The water molecules of the adlayer form hexagonal rings, which are experimentally and theoretically observed for water adsorbed at (111) surfaces of several fcc metals [139, 147, 148]. The vacancies in the honeycomb-like structure are locally occupied resulting in an average coverage of 70% of all top sites.

The plane-averaged density of the water molecules as function of the distance from the surface is displayed in Figure 3.15(c). The sharp peak at 2.36 Å indicates the adlayer followed by a ∼3Å thick region of density depletion. A second peak is found at about 5 Å showing that the metallic substrate also dictates the water structure in the second adsorption layer. Compared to bulk water, the water molecules in these regions are expected to be massively overstructured. Further away from the surface, the density oscillations decay rapidly. For distances larger than 10 Å, the density

Table 3.4.: Dipole moments μ in Debye for $12H_2O$ on Pt(111). $\Delta|\mu|$ is the deviation of the dipole moments comparing IC-QM/MM and QM/MM. The molecule numbers correspond to the numbering in Figure 3.14(a).

		IC-QM/MM structure		QM/MM structure	
molecule nr.	$\lvert\mu\rvert_{\mathrm{IC-QM/MM}}$	$\lvert\mu\rvert_{\mathrm{QM/MM}}$	$\Delta\lvert\mu\rvert$	$\lvert\mu\rvert_{\mathrm{QM/MM}}$	$\Delta\lvert\mu\rvert$
1	2.52	2.44	0.08	2.47	0.05
2	2.64	2.49	0.15	2.49	0.15
3	2.45	2.38	0.07	2.14	0.31
4	2.71	2.55	0.16	2.48	0.23
5	2.67	2.53	0.14	2.52	0.15
6	2.54	2.46	0.08	2.40	0.14
7	2.42	2.29	0.13	2.18	0.24
8	2.52	2.39	0.13	2.34	0.18
9	2.75	2.64	0.11	2.53	0.22
10	2.01	1.98	0.03	1.97	0.04
11	2.30	2.26	0.04	2.25	0.05
12	2.25	2.17	0.08	2.17	0.08

corresponds to the one of bulk water. At the liquid-vacuum interface, the density finally decreases to zero.

For further discussion of structural and electronic properties, the water film was split in four separate layers guided by the oscillations in the density profile, see Figure 3.15(c). Radial distribution functions (RDFs) for each layer are reported for oxygen-oxygen (O-O) and oxygen-hydrogen (O-H) in Figure 3.16(a) and (b). The curves for the second and third layer resemble RDFs of bulk liquid water obtained with DFT-PBE in previous studies [143, 149]. The peaks for the water-metal and water-vacuum layers are located at the same distance. However, the peak heights of the fourth layer are significantly smaller due to the reduced coordination at the water-vacuum interface. Even though the coordination is also reduced for the adlayer along the z-axis, the peak heights at 2.7 Å (O-O) and 1.7 Å (O-H) are comparable to bulk water. For larger distances, the RDFs decay to values smaller than 1 and show several additional peaks. This is in agreement with the plane-averaged density profile predicting an overstructuring in the adlayer which compensates the lacking coordination in the third dimension for small, but not for large distances.

First insights in the nature of hydrogen bonding at the metallic interface are given by the plane-averaged distributions of O and H atoms, see Figure 3.15(d). The onset of the hydrogen distribution is at shorter distances thanks to the reorientation induced by the polarization of the image charges, i.e. some H atoms are pointing towards the substrate, which was already observed for the water clusters. The largest fraction of

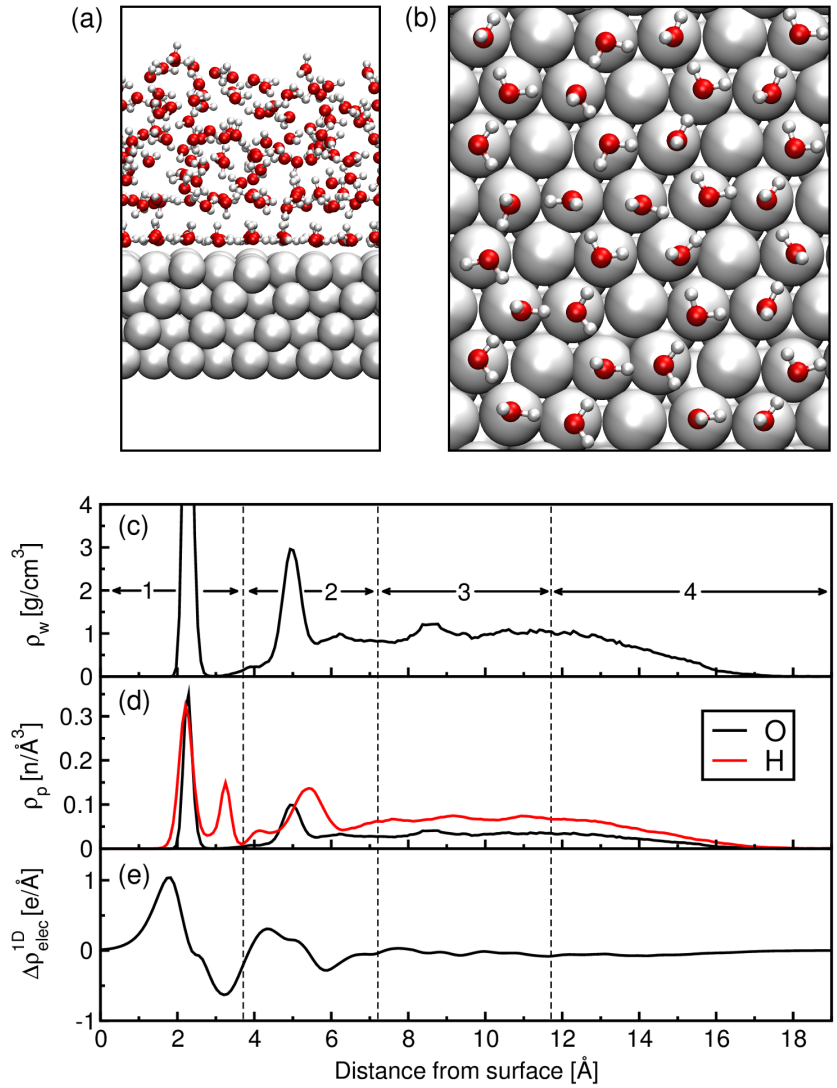

Figure 3.15.: Structural and electronic properties obtained for a water film on Pt(111). (a) Typical snapshot of the water/Pt(111) simulation cell. (b) Snapshot of the water adlayer. (c) Plane-averaged water density ρ_w along the surface normal. The arrows indicate the division of the water film in 4 different layers. (d) Plane-averaged particle density ρ_p for O and H atoms. (e) Plane-integrated electronic density difference $\Delta\rho^{1D}_{elec}$ between water and Pt(111).

the hydrogens of the adlayer are located in the plane of the oxygen atoms indicated by the large peak at 2.36 Å. The small H atom peak at 3.2 Å, located in the region of density depletion, implies that a fraction of the water molecules in the adlayer forms H-bonds with the acceptors of the second layer. This is confirmed by more detailed analysis using a standard criterion for defining an H-bond [150,151]. Almost half of the water molecules of the adlayer donate one of their H atoms to the bulk. The second H atom functions as H-bond donor in the plane of the adlayer. For further investigation of the H-bond network, the average number of H-bonds for each molecule (accepted

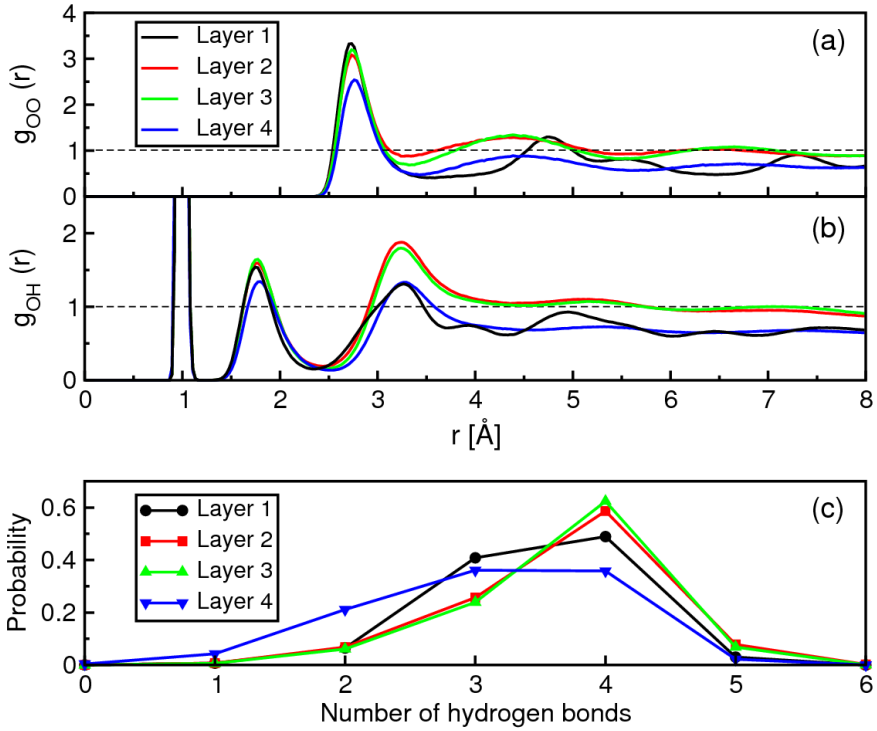

Figure 3.16.: Structural properties for the four different layers of the water film defined in Figure 3.15(c). (a) Oxygen-oxygen and (b) oxygen-hydrogen RDFs. (c) Distribution of the total number of H-bonds per molecule.

and donated) was calculated and the distribution for each layer has been plotted in Figure 3.16(c). From the results we can see that most of the molecules of the second and third layer have four H-bonds. At the liquid/metal interface, the probability of the formation of four H-bonds is significantly reduced, while the probability of forming only three H-bonds increases. From that we can conclude that the geometry of the metal surface dictates hydrogen bonding patterns that are incommensurate with the tetrahedral structure characteristic for bulk water. This leads to unfavorable interactions with the adjacent liquid resulting in a region of density depletion.

Simulations of a similar system within a purely classical frame using the same empirical interaction potential and a classical version of the IC approach are reported in Ref. [105]. We obtain very similar results for the profile of the water density and the structure of the adlayer. However, in the purely classical model, more vacancies are occupied resulting in a 15% larger coverage. The average density of hydrogens donated to the bulk is approximately five times larger with our method. This suggests that in our IC-QM/MM description the adlayer is less hydrophobic than predicted by the purely classical model. The latter implies that the interactions between the water molecules predicted by the DFT-PBE potential are stronger than for the classical potential (SPC/E), which is confirmed by structural and dynamical properties such as RDFs or diffusion coefficients [149, 152, 153] and thermodynamic properties [154].

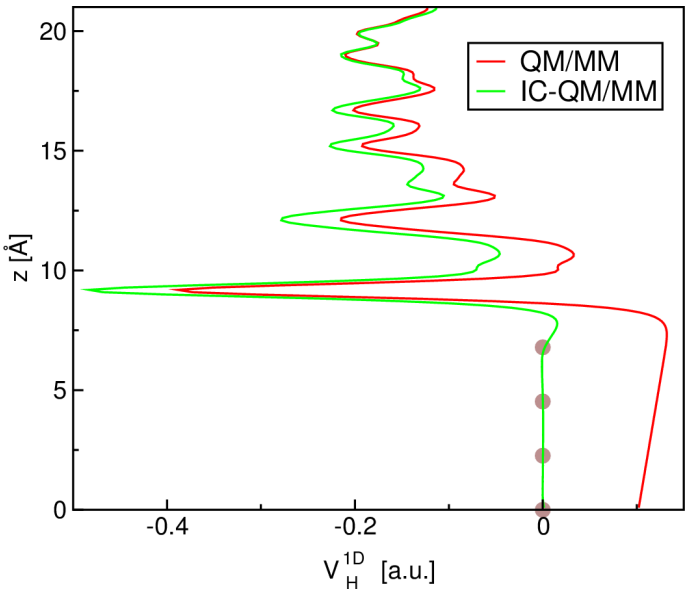

Figure 3.17.: Plane-averaged electrostatic potential $V_{\mathrm{H}}^{1\mathrm{D}}$ of 151 H_2O on Pt(111) for a particular snapshot of the MD simulation. Brown spheres indicate the position of the metal layers. $V_{\mathrm{H}}^{1\mathrm{D}}$ has been averaged along the surface normal (z-direction).

Comparing the structural properties of the water/metal interface predicted by IC-QM/MM to experiments and *ab initio* calculations, we find that our model reproduces some crucial features of the $\sqrt{39}$ structure. The water molecules form H-bonds predominately in the plane of the adlayer and nonbonded hydrogens point down to the surface. Hexagonal rearrangements are also structural motives occurring in the $\sqrt{39}$ structure. However, the formation of five- and seven-membered rings is not captured.

IC-QM/MM allows also some insight in the electronic rearrangements upon adsorption of the water film. The plane-integrated electronic density difference is displayed in Figure 3.15(e). Electron accumulation is observed below and depletion above the adlayer. Compared to the adsorption of a single nonpolar molecule like benzene, the polarization is orders of magnitude larger. The electronic polarization extends also into the second layer, but decreases to zero in the third layer. The average dipole moment in the bulk-like layers is ~ 3.0 Debye and about 0.2 Debye smaller at the liquid/vacuum interface due to a reduced number of H-bonds. Even though the latter is true at the metal/liquid interface, the average dipole moment is also 3.0 Debye which must be attributed to polarization effects.

The observed structural and electronic properties are strongly affected by the presence of the IC interactions with the QM molecules. This is easily verified by switching off these interactions and following the MD trajectory as generated by the bare QM/MM scheme. As a result, the first layer of water molecules in contact with the Pt surface is destabilized. The ordered arrangement is disrupted in a few ps and some molecules dissociate. This unphysical phenomenon is most probably a consequence

of the strong distortion forces arising by suddenly switching off the image charges. As shown in Figure 3.17, the IC contributions significantly change the electrostatic potential over a range of more than 10 Å above the surface.

3.5. Conclusions

In this chapter, an image charge augmented QM/MM method has been introduced to enable QM/MM simulations of molecules adsorbed on metallic surfaces. With IC-QM/MM, important polarization effects are included explicitly. The method has been evaluated using a few simple test systems such as benzene, nitrobenzene, thymine and guanine on Au(111). It has been demonstrated that our model reproduces the essential physics, i.e. the potential within the metal is screened and the induced charges display the expected distribution in the substrate layers. We could show that the IC-QM/MM method substantially strengthens the interaction between molecules and metal and can reproduce at least partially the electrostatic response of the molecular charge density to the induced charges. Close to the surface, the rearrangement of the electronic structure is not sufficiently captured by a pure polarization model since the molecular and metallic charge densities are not separated. However, IC-QM/MM is expected to model the long-range interactions correctly.

The method has been further evaluated with small water clusters adsorbed at Pt(111). Interaction and adsorption energies agree well with full DFT results set as benchmark. Pronounced structural changes attributed to polarization effects are observed for the largest cluster. By comparing to experimental data, we could show that the IC-QM/MM structure is reasonable.

We applied our method to perform an MD simulation of a liquid water/Pt(111) interface. Structural analysis of the water film and especially the adlayer, reveals that the water molecules bind strongly to the surface and that the hydrogen bonding is frustrated by the geometry of the surface. The water-metal interface is thus hydrophobic, but more hydrophilic than predicted by purely classical simulations. As discussed in Ref. [105], the reorganization in the adlayer occurs at timescales larger than nanoseconds. Obtaining dynamical properties at these timescales with *ab initio* MD is out of scope, but possible to address with the IC-QM/MM approach. IC-QM/MM is potentially a valuable tool to study these interfaces with large-scale MD and to allow implications on electrochemical properties of metal electrodes. Successful application of the classical equivalent of the IC approach in electrochemistry [107–109] also suggests that the method holds promise especially for electrochemical systems.

4. Wetting of water on hexagonal boron nitride@Rh(111): A QM/MM model based on atomic charges derived for nano-structured substrates [155]

The wetting of water on corrugated and flat hexagonal boron nitride (h-BN) monolayers on Rh(111) is studied within a hybrid quantum mechanics/molecular mechanics (QM/MM) approach. Water is treated by QM methods, whereas the interactions between liquid and substrate are described at the MM level. The electrostatic properties of the substrate are reproduced by assigning specifically generated partial charges to each MM atom. We propose a method to determine restrained electrostatic potential (RESP) charges that can be applied to periodic systems. The approach is based on the Gaussian and plane waves algorithm and allows an easy tuning of charges for nano-structured substrates. We have successfully applied it to reproduce the electrostatic potential of the corrugated and flat h-BN/Rh(111) known as nanomesh. Molecular dynamics simulations of water films in contact with these substrates are performed and structural and dynamic properties of the interfaces are analyzed. Based on this analysis and on the interaction energies between water film and substrate, we found that the corrugated nanomesh is slightly more hydrophilic. On a macroscopic scale, a smaller contact angle is predicted for a droplet on the corrugated surface.

4.1. Introduction

Wetting behavior of liquid water on surfaces is a long-standing research field with major importance for technological applications such as sensors and coatings [156, 157]. Deeper understanding of the wetting at functional surfaces is especially important for fluids confined in nanoscale structures. Controlling the wettability at the nanoscale potentially enables the development of smart surfaces and nano-based devices. In this respect, versatile templates like single layers of hexagonal boron nitride (h-BN) on metal surfaces are of particular interest. Monolayers of h-BN have been epitaxially grown on hot metallic surfaces by chemical vapor deposition of borazine [158] molecules. The structural and electronic properties of BN can be tuned by the choice of the metal [159, 160]. h-BN structures have been prepared on a range of transi-

tion metals, such as Rh(111), [158, 161] Ni(111), [162–164] Cu(111), [162, 165, 166] Ir(111) [167, 168] and several others [169–173].

On Rh(111), a corrugated superstructure is formed with a periodicity of 3.2 nm induced by the lattice mismatch between BN and Rh [161, 174, 175]. The unit cell of this so-called nanomesh is given by 13×13 BN units on 12×12 Rh(111). The nanomesh is characterized by a regular arrangement of 2 nm "pores" separated by regions less tightly bound to Rh(111) called "wire". The corrugation causes a modulation in the electrostatic potential (ESP) above the BN layer, which is thought to be responsible for trapping molecules in the pores [176]. Preferential adsorption in the pore regions has been studied for organic molecules, [161, 177, 178] noble gas atoms [179] and to great extent for small water clusters [180–184]. The mesh structure is very robust, stable in gas phase, [185] survives immersion in liquids [161, 186] and withstands temperatures up to 1000 K [161]. By intercalation of atomic H, which strongly binds to the Rh surface, the interaction between nitrogen and Rh is weakened and the *h*-BN layer becomes much flatter. However, this modification can be easily reversed by desorption of hydrogen, thus recovering the original nanomesh [187]. By reversible intercalation and removal of H atoms, switching between corrugated and flat structure is possible. Since the corrugation determines the functionality of the surface, the interactions between molecules and substrate can be reversibly altered. This is also interesting in the context of wetting in order to control the hydrophilicity of the surface.

In this work, the wetting of liquid water on the corrugated and the H-intercalated nanomesh is studied. The two substrates are in the following referred to as nm and nm-H, respectively. A full quantum mechanical description of the wetting by density functional theory (DFT) is not affordable because of computational limitations. Instead, a computationally less expensive hybrid scheme combining quantum mechanics (QM) with molecular mechanics (MM) is used [1]. In our QM/MM setup, the liquid water is described by DFT, whereas the substrate and the interactions between substrate and molecules are treated at the MM level of theory. A similar setup has been used in our previous work and proved to be successful for adsorbate-metal systems [90]. In order to properly describe the electrostatic interactions between substrate and liquid, the very specific modulated features of the ESP above the nanomesh have to be reproduced. These features have been extensively characterized by DFT calculations of the electronic structure of the nanomesh [160,183]. MM force fields typically rely on partial atomic charges to model electrostatic interactions. A popular approach is to apply a fitting procedure in order to determine the charges that optimally reproduce a given ESP. Often, restraints and constraints are also implemented to guarantee physically meaningful values of the resulting charges, which are then called RESP charges [188]. This approach is widely used for nonperiodic molecular systems. Fitting RESP charges for periodic systems leads to additional difficulties due to the long-range nature of the ESP. Periodic RESP schemes were derived and successfully applied to molecular crystalline systems like metal organic frameworks [189, 190]. We

present in this work a novel adaptation and extension of the RESP charge fitting procedure for periodic systems that are slab-like. Our approach is based on a mixed Gaussian and plane wave (GPW) formalism [44,118] and hence naturally dealing with periodic ESPs and additionally circumventing problems related to their ill-defined reference state.

In the next section, the theory and implementation of our periodic RESP model are described. Adjustable parameters of the QM/MM setup are discussed and validated. RESP charges are derived for nm and nm-H and the accuracy of the fitted potential is assessed. Molecular dynamics (MD) simulations of liquid water at nm and nm-H interfaces are performed using the previously described QM/MM setup. Structural and dynamic properties are discussed for insight in the wetting at molecular scale. We address interaction energies and the contact angle of a water droplet on these substrates to evaluate and compare the hydrophilicity in macroscopic terms.

4.2. Periodic RESP fitting

4.2.1. Theory

For obtaining RESP charges, a quantum chemical calculation is performed to generate a reference ESP V_{QM}. A set of atomic point charges $\{q_a\}$ is derived such that the difference between V_{QM} and the potential V_{RESP} generated by $\{q_a\}$ is minimized. This minimization is performed in a least squares fitting procedure at defined grid points $\mathbf{r}_k$. The residual that needs to be minimized is given by

$$R_{\mathrm{esp}} = \frac{1}{N}\sum_{k}^{N}\left(V_{\mathrm{QM}}(\mathbf{r}_k) - V_{\mathrm{RESP}}(\mathbf{r}_k)\right)^2, \tag{4.1}$$

where N is the total number of grid points included in the fitting procedure. The optimal choice of $\{\mathbf{r}_k\}$ depends on the system and is discussed below. In our approach, the point charges are represented as Gaussian functions with fixed width α. The resulting charge distribution ρ is defined as

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_a q_a g_a(\mathbf{r}, \mathbf{R}_a) \\ &= \sum_a q_a \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp(-\alpha|\mathbf{r} - \mathbf{R}_a|^2),\end{aligned} \tag{4.2}$$

where $\mathbf{R}_a$ is the position of atom a and g_a is the normalized spherical Gaussian function centered at a. The potential V_{RESP} generated by ρ and given by

$$V_{\mathrm{RESP}}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' = \sum_a q_a \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \tag{4.3}$$

is inserted in Equation 4.1. When calculating RESP charges for periodic systems, periodic boundary conditions (PBC) have to be employed for the calculation of V_{RESP}. Our approach is based on the GPW formalism. Within the GPW approach, the RESP charge distribution ρ can be represented on a regular grid. Fast Fourier transforms (FFT) are used to change the representation from real space $\rho(\mathbf{r})$ to reciprocal space $\rho(\mathbf{G})$, where $\mathbf{G}$ is the reciprocal lattice vector. Since the Poisson equation is solved in reciprocal space, the obtained potential $V_{\mathrm{RESP}}(\mathbf{G})$ is naturally periodic. An inverse FFT is used to transfer the potential back to real space.

To avoid physically meaningless atomic charges, restraints R_{rest} and constraints R_{const} are introduced in the fitting procedure. The total residual is consequently

$$R = R_{\mathrm{esp}} + R_{\mathrm{rest}} + R_{\mathrm{const}}. \tag{4.4}$$

To address restraints, a harmonic penalty function is chosen,

$$R_{\mathrm{rest}} = \beta \sum_j (q_j - t_j)^2, \tag{4.5}$$

where t_j is the target charge and β a scaling factor determining the strength of the restraint. Various constraints can be employed. Usually, the sum of all fitted charges is constrained to the total charge q_{tot} of the system,

$$R_{\mathrm{const}} = \lambda \sum_j (q_j - q_{\mathrm{tot}}), \tag{4.6}$$

where λ is the Lagrange multiplier. Additional constraints can be employed, for example, to ensure that specific atoms carry the same charge.

To obtain the set of atomic charges $\{q_a\}$, the residual R needs to be minimized with respect to q_a. Starting with R_{esp}, the minimization yields a linear set of equations,

$$\begin{aligned}
\frac{\partial R_{\mathrm{resp}}}{\partial q_a} &= \frac{2}{N} \sum_k (-V_{\mathrm{QM}}(\mathbf{r}_k) + V_{\mathrm{RESP}}(\mathbf{r}_k)) \frac{\partial V_{\mathrm{RESP}}(\mathbf{r}_k)}{\partial q_a} \\
&= \frac{2}{N} \sum_k (-V_{\mathrm{QM}}(\mathbf{r}_k) + V_{\mathrm{RESP}}(\mathbf{r}_k)) \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}' \\
&= \frac{2}{N} \sum_k -V_{\mathrm{QM}}(\mathbf{r}_k) \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}' + \sum_b q_b \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}' \int \frac{g_b(\mathbf{r}', \mathbf{R}_b)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}' \\
&= -C_a + \sum_b q_b T_{ab} = 0.
\end{aligned} \tag{4.7}$$

The vector $\mathbf{C}$ is given by

$$C_a = \frac{2}{N} \sum_k V_{\mathrm{QM}}(\mathbf{r}_k) \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}' \tag{4.8}$$

and matrix $\mathbf{T}$ by

$$T_{ab} = \frac{2}{N} \sum_k \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}' \int \frac{g_b(\mathbf{r}', \mathbf{R}_b)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}'. \tag{4.9}$$

Including also the residuals for the restraints and constraints, the minimization condition modifies to

$$\frac{\partial R}{\partial q_a} = -C_a + \sum_b q_b T_{ab} + 2\beta(q_a - t_a) + \lambda = 0. \tag{4.10}$$

Storing the unknowns $\{q_a\}$ and λ in vector $\mathbf{Q}$, their solutions is obtained by solving the linear set of equations

$$\mathbf{AQ} = \mathbf{B}. \tag{4.11}$$

If n is the number of atoms, the matrix elements A_{ab} are defined for $a, b \leq n$ in terms of the Kronecker delta δ_{ab},

$$A_{ab} = T_{ab} + 2\beta\delta_{ab} \tag{4.12}$$

indicating that 2β is only added to the diagonal elements of matrix $\mathbf{T}$. The other elements are in this case 1 except for $A_{n+1,n+1}$. Depending on the number and type of constraints, matrix $\mathbf{A}$ is modified accordingly. For $a \leq n$, vector $\mathbf{B}$ is given by

$$B_a = C_a + 2\beta t_a. \tag{4.13}$$

The element B_{n+1} is here equal to the total charge of the system. In matrix representation this can be summarized as

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ A_{n1} & \cdots & A_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ \vdots \\ q_n \\ \lambda \end{pmatrix} = \begin{pmatrix} B_1 \\ \vdots \\ B_n \\ q_{\text{tot}} \end{pmatrix} \tag{4.14}$$

and solved by a conventional Gaussian elimination scheme or a conjugate gradient procedure.

4.2.2. Implementation

The periodic RESP fitting based on the GPW approach is part of the CP2K program package [121, 191]. In order to obtain a periodic solution of V_{RESP}, PBC have to be applied for all parts of the fitting procedure. Periodic solutions are required for vector C_a and matrix T_{ab}, which can be reformulated in terms of the potential $v_a(\mathbf{r}_k)$,

$$C_a = \frac{2}{N} \sum_k V_{\text{QM}}(\mathbf{r}_k) v_a(\mathbf{r}_k) \tag{4.15}$$

Loop over all a atoms

Collocate $g_a(\mathbf{r})$ on the real space grid

Transfer to reciprocal space: $g_a(\mathbf{r}) \xrightarrow{\text{FFT}} g_a(\mathbf{G})$

Solve Poisson equation: $g_a(\mathbf{G}) \longrightarrow v_a(\mathbf{G})$

Transfer back to real space: $v_a(\mathbf{G}) \xrightarrow{\text{FFT}^{-1}} v_a(\mathbf{r})$

Loop over all atoms b

T: Calculate product $v_a(\mathbf{r}_k)v_b(\mathbf{r}_k)$ at grid points $\mathbf{r}_k$

End b loop

C: Calculate product $V_{\text{QM}}(\mathbf{r}_k)v_a(\mathbf{r}_k)$ at grid points $\mathbf{r}_k$

End a loop

Figure 4.1.: Pseudocode for calculating matrix **T** and vector **C**.

and

$$T_{ab} = \frac{2}{N} \sum_k v_a(\mathbf{r}_k) v_b(\mathbf{r}_k). \tag{4.16}$$

The potential $v_a(\mathbf{r}_k)$ is generated by one single Gaussian charge at atom a and given by

$$v_a(\mathbf{r}_k) = \int \frac{g_a(\mathbf{r}', \mathbf{R}_a)}{|\mathbf{r}' - \mathbf{r}_k|} d\mathbf{r}'. \tag{4.17}$$

The procedure to obtain periodic solutions for $v_a(\mathbf{r}_k)$ and subsequently T_{ab} and C_a is shown in Figure 4.1. The potential $v_a(\mathbf{r}_k)$ is transferred to real space by an inverse FFT step once the Poisson equation has been solved in reciprocal space. The matrix and vector elements are then computed on the real space grid for the set of pre-defined grid points $\{\mathbf{r}_k\}$. This computation is performed for all n atoms. As already discussed in Section 3.2.2, the single Gaussian charges $g_a(\mathbf{r})$ are in fact located at different positions on the grid, but generate the same potential v_a. Theoretically, v_a could be calculated for one reference atom a_{ref} since it holds that $v_a(\mathbf{r}_k) = v_{a_{ref}}(\mathbf{r}'_k)$ where $\mathbf{r}'_k = \mathbf{r}_k - \mathbf{R}_a + \mathbf{R}_{a_{ref}}$. This would massively reduce the computational effort by decreasing the number of FFTs from $2n$ to 2. Since we use a grid-based approach, the atom positions might not coincide with any of the grid points. Hence, the equality $v_a(\mathbf{r}_k) = v_{a_{ref}}(\mathbf{r}'_k)$ is only an approximation where the accuracy depends on the coarseness of the grid. RESP fitting is usually a post-processing step and performed after a single-point calculation. Thus, trading accuracy for speed is not crucial here and calculating v_a explicitly for all n atoms is the preferred option.

Our implementation of periodic RESP fitting utilizes the massive parallelization procedures already present in CP2K and enables an efficient computation of RESP charges also for large systems.

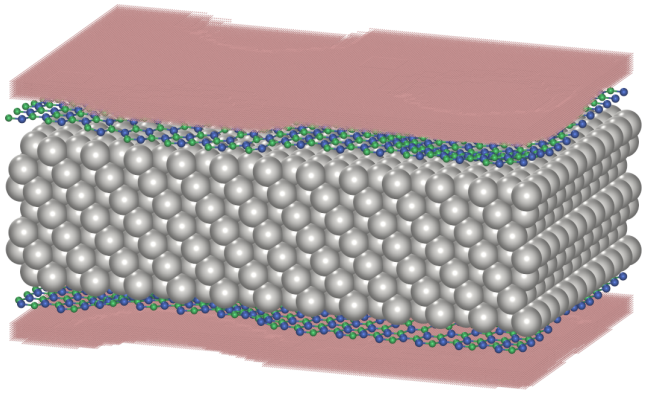

Figure 4.2.: Grid points sampled in the range of 2-4 Å above the surface for the symmetric corrugated *h*-BN/Rh(111) slab. Color code: B green, N blue , Rh gray, grid points pink.

In calculations of periodic ESPs, the arbitrariness of the reference state needs to be considered. Campañá et al. [189] pointed out that reference and fitted potential can suffer from different constant offsets, which they addressed by modifying the residual R accordingly. However, in our approach V_{QM} and V_{RESP} are computed in the same way using the GPW scheme. Hence they have the same constant offset which cancels out when evaluating R_{esp}.

4.2.3. Sampling of fit points

A set of grid points is chosen where the atomic charges are fitted to correctly reproduce the reference potential. In molecular calculations, the grid points that are included in the fitting are usually chosen to lie outside the van der Waals region where intermolecular interactions have to be described correctly. In addition, regions of high electron density, e.g., within the van der Waals regions, cannot be modeled accurately by atomic charges anyway. For slab-like systems, the ESP has to be accurately reproduced over the surface which is especially important when studying adsorption processes. A good choice is to sample the ESP in a specified range of, e.g., 2-4 Å above the surface. In case of corrugated substrates like nm, our sampling is flexible enough to follow the corrugation in a defined volume above the BN layer as displayed in Figure 4.2.

4.3. Computational details

4.3.1. Charge fitting

The superstructure of nm is represented by a hexagonal arrangement of 13×13 *h*-BN on 12×12 replica of the Rh unit cell. For nm-H, 12×12 H atoms are intercalated between the *h*-BN and the topmost Rh(111) layer. We model nm by a slab of seven

Rh layers terminated on both sides by one h-BN monolayer. PBC are applied in all calculations. Spurious interactions with periodic images are avoided by adding a vacuum of at least 15 Å in the direction perpendicular to the surface.

All calculations are carried out with the CP2K program package [121, 191]. Structure optimization and calculation of V_{QM} for nm and nm-H are performed at the Kohn-Sham DFT level within the GPW approach [44, 118]. Single-ζ valence plus polarization (SZVP) basis sets are used to describe the valence electrons of Rh, whereas basis sets of at least double-ζ quality (DZVP) are used for the other elements [122]. Norm-conserving pseudo-potentials [123–125] are employed to approximate the interactions between valence electrons and atomic cores. The exchange and correlation potential is given by the revised Perdew-Burke-Ernzerhof (revPBE) [192] functional. The latter is used in combination with Grimme's D3 correction [34] to account for long-range dispersion interactions. The cutoff for the plane wave (PW) expansion of the auxiliary density is set to 500 Ry. Fermi-Dirac smearing [193] at an electronic temperature of 300 K and Broyden density mixing [194] are employed to guarantee a smooth convergence.

For the fitting procedure of the RESP charges, the total charge of the system is constrained to zero and atoms of pre-defined groups are constrained to carry identical charges. Moreover, restraints are employed as described below. To evaluate the quality of the fit, V_{RESP} is compared to the reference DFT potential by means of the root-mean square (RMS) error,

$$\mathrm{RMS} = \sqrt{\frac{\sum_k^N \left(V_{\mathrm{QM}}(\mathbf{r}_k) - V_{\mathrm{RESP}}(\mathbf{r}_k)\right)^2}{N}} \tag{4.18}$$

and relative root-mean square (RRMS) error

$$\mathrm{RRMS} = \sqrt{\frac{\sum_k^N \left(V_{\mathrm{QM}}(\mathbf{r}_k) - V_{\mathrm{RESP}}(\mathbf{r}_k)\right)^2}{\sum_k^N V_{\mathrm{QM}}(\mathbf{r}_k)^2}}. \tag{4.19}$$

4.3.2. MD simulations

QM/MM has been used to perform MD simulations of liquid water at the nm and nm-H interface. For the MD simulations, the hexagonal h-BN/Rh(111) slabs have been extended to orthorhombic slabs with 13×26 h-BN and 12×24 Rh. Both slabs are surmounted with a 15 Å thick water layer consisting of 833 water molecules. A vacuum space of 25 Å is added to prevent interactions with periodic images along the [111] direction. The water layer is treated at the DFT level of theory, whereas the interactions between water and the nanomesh are computed at the MM level. The positions of the substrate atoms are kept fixed in their DFT optimized structure. For the QM part DZVP basis sets, the PBE functional [20] and a PW cutoff of 300 Ry are used.

The electrostatic interactions between water layer and nanomesh are modeled by the RESP charges given in Table 4.2. The electrostatic coupling procedures implemented in the QM/MM [119, 120] part of CP2K are employed to correctly treat the electrostatics with PBC. Note that QM and MM box have been set to the same size in order to avoid computationally expensive decoupling between QM images. To reproduce dispersion and Pauli repulsion between water and nanomesh, an empirical potential of the Lennard-Jones (LJ) type is used. The B-O and N-O interaction parameters are $\sigma_{\text{B-O}} = 3.31\,\text{Å}$, $\sigma_{\text{N-O}} = 3.26\,\text{Å}$, $\epsilon_{\text{B-O}} = 0.52\,\text{kJ/mol}$ and $\epsilon_{\text{N-O}} = 0.63\,\text{kJ/mol}$. The latter were derived from the B-B and N-N interaction parameters given in Ref. [195, 196] and the O-O parameters of the SPC/Fw water model [197].

All MD simulations are performed within the canonical (NVT) ensemble using a time step of 0.5 fs. Before starting the actual simulations, all systems have been pre-equilibrated solely by MM for at least 1 ns using the SPC/Fw water potential and temperature annealing up to 300 K. The QM/MM simulations are run at a temperature of 330 K and equilibrated for at least 10 ps. The last 20 ps of the simulations have been used for analysis.

To obtain interaction energies and properties related to the contact angle, a DFT-based MD simulation of a pure 15 Å thick water slab is performed with the same number of molecules. The interaction energies E_{int} between nm and water film are computed per unit area according to

$$E_{\text{int}} = \frac{E_{\text{nm+H}_2\text{O}} - (E_{\text{nm}} + E^{\text{slab}}_{\text{H}_2\text{O}})}{A}, \tag{4.20}$$

where $E_{\text{nm+H}_2\text{O}}$ denotes the averaged potential energy of nm surmounted with the water layer. E_{nm} is the energy of the substrate only and $E^{\text{slab}}_{\text{H}_2\text{O}}$ the averaged energy of the pure water slab. A is the area defined by the horizontal dimensions of the slab systems.

The contact angle θ of a water droplet on a surface can be estimated from interface energies according to Young's formula, [198–200]

$$\cos\theta = \frac{\gamma_{\text{SV}} - \gamma_{\text{SL}}}{\gamma_{\text{LV}}}, \tag{4.21}$$

where γ_{LV}, γ_{SL} and γ_{SV} represent the energies of the liquid-vacuum, solid-liquid and the solid-vacuum interfaces, respectively. For the nanomesh, γ_{SV} is not accessible since the corresponding bulk-phase of nm cannot be properly defined. However, the difference of $\cos\theta$ between nm and nm-H can be computed assuming that γ_{SV} is identical for both systems. Following the argumentation in Ref. [201], we estimate γ_{SL} by the interaction energy per unit area. For the liquid-vacuum interface, γ_{LV} is approximated by

$$\gamma_{\text{LV}} = \frac{E^{\text{slab}}_{\text{H}_2\text{O}} - \frac{n_{\text{slab}}}{n_{\text{bulk}}} E^{\text{bulk}}_{\text{H}_2\text{O}}}{2A}, \tag{4.22}$$

Table 4.1.: RESP charges for boron and nitrogen atoms of the wire region of the pristine h-BN/Rh(111) obtained with different strengths β of the harmonic restraints. The charges have been restrained to the target values of 0.255 and −0.255, respectively

		β		
	target value	0.01	0.001	0.0001
B4	0.255	0.239	0.211	0.201
N4	−0.255	−0.272	−0.302	−0.311
RRMS		0.61	0.14	0.02

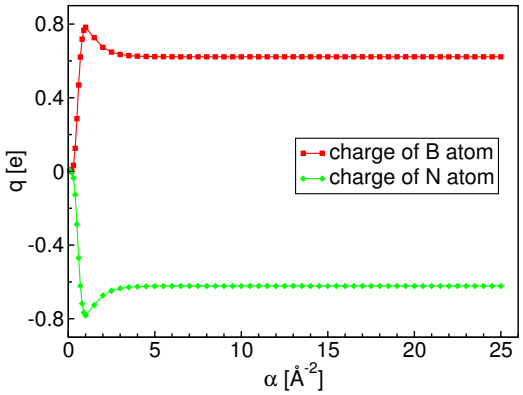

Figure 4.3.: Dependence of the RESP charge q on the Gaussian width α. Charges fitted for a single h-BN sheet.

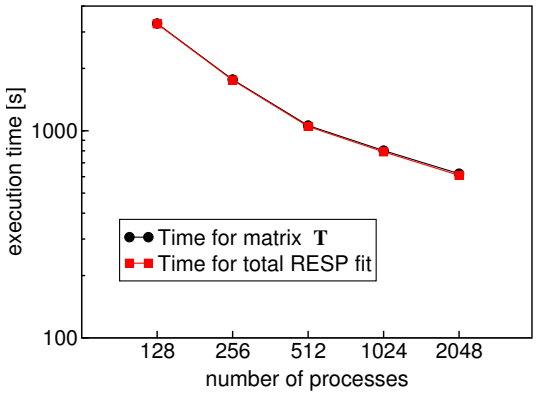

Figure 4.4.: Execution time for RESP fitting and for calculating matrix **T** on a Cray XC30 system.

where $E^{\text{bulk}}_{H_2O}$ is the averaged potential energy for bulk water and n_{slab} and n_{bulk} are the number of water molecules in the slab and bulk, respectively. To assess $E^{\text{bulk}}_{H_2O}$, bulk water is simulated at the DFT level using a cubic box with 128 water molecules.

4.4. Tests and results

4.4.1. Parameter validation and performance

The RESP approach relies on restraining charges during the fitting process, see Equation 4.5. Target values for the charges and the strength of the restraint β are tunable parameters. The larger β, the less flexible is the fitting leading to larger RRMS errors. We found that values of $\beta > 10^{-4}$ drastically increase the RRMS value, see Table 4.1. Therefore, β was always set to 10^{-4} yielding physical charges and an accurately fitted potential. The target values were partly estimated from unrestrained ESP charges and density derived Blöchl charges [202].

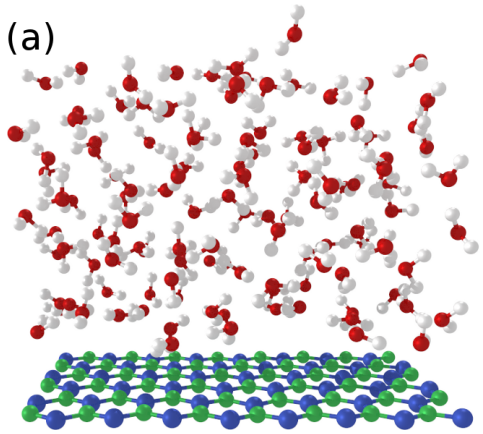

Figure 4.5.: Water slab on a free-standing BN layer. The setup of the calculations corresponds to the one described in Section 4.3. (a) Snapshot of the simulation cell containing 120 water molecules and (b) interaction energies E_{int} between water film and BN layer.

The Gaussian width α of the charge distribution ρ is another adjustable parameter in our approach, see Equation 4.2. To evaluate the influence of α on the fitted charges q, test runs of a single h-BN sheet have been performed. The dependency of the magnitude of q on α is reported in Figure 4.3. The charges do not depend on the Gaussian width for values of α larger than 3.0 Å^{-2}. However, the fitting procedure becomes unreliable for smaller values. Small values of α model too broad Gaussian functions describing a uniform charge distribution rather than localized charges, which necessarily leads to wrong results.

The LJ parameters modeling the dispersion were not specifically optimized for nm and nm-H. To assess the error introduced by inaccurate dispersion interactions, MD simulations of a water slab in contact with a free-standing BN layer have been performed, see Figure 4.5(a). MD simulations have been carried out with full DFT meaning that the whole system is treated by DFT and with QM/MM treating interactions between water and BN at the MM level. For this system, electrostatic interactions at the interface can be neglected. The MM-based interactions between BN and water have been assumed to be of the LJ type. Comparing the interaction energies obtained by full DFT simulations and QM/MM provides a direct measure of the quality of the LJ parameters given in Section 4.3.2. The interaction energy based on the LJ parameters is larger by a factor of 1.25 indicating that our QM/MM setup is slightly overestimating the dispersive interactions, see Figure 4.5(b).

The performance of the periodic RESP fitting is evaluated for nm and displayed in Figure 4.4. The calculation of matrix $\mathbf{T}$ is by far the computationally most expensive step and dominates the timings. Using an orthorhombic model for nm, the potential $v_a(\mathbf{r}, \mathbf{R}_a)$ has to be evaluated 3368 times, which sums up to 6736 FFTs. However, the time needed for the RESP fitting is order of magnitudes smaller than the time needed to converged the wavefunction for such a large system. The parallel efficiency

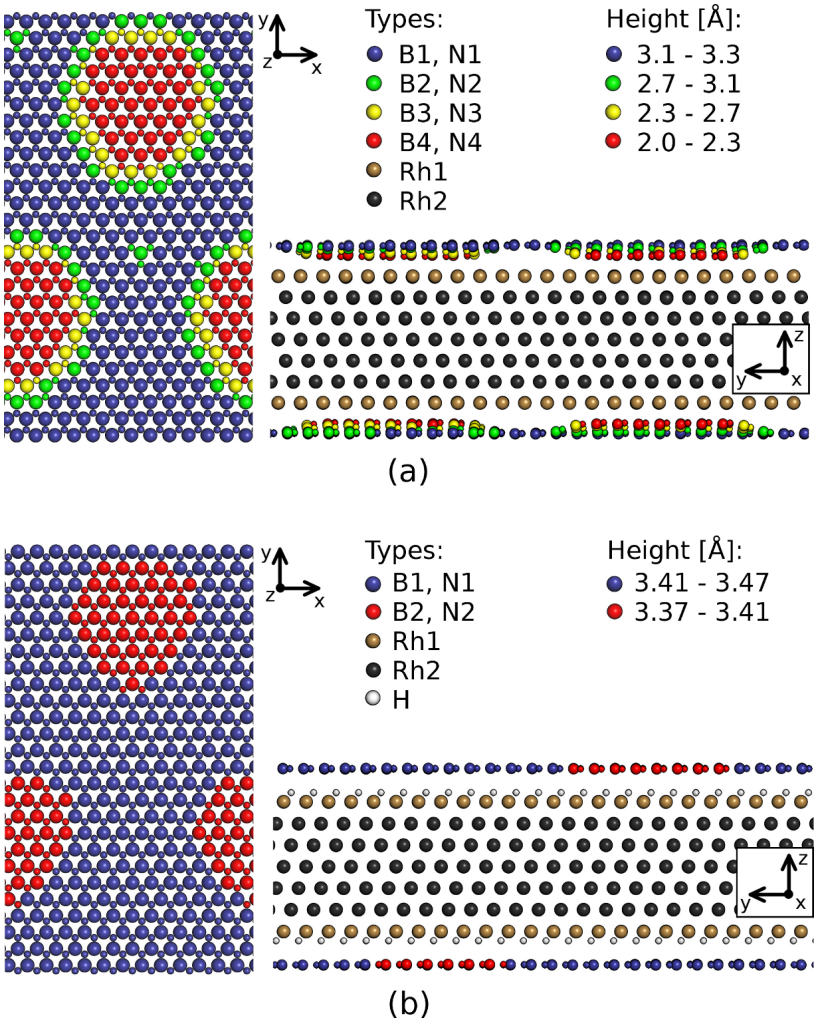

Figure 4.6.: Assignment of the atom types for (a) nm and (b) nm-H. The height of B and N atoms is measured relative to the average height of the topmost Rh layer (Rh1). Top view of the BN layer (left) and side view of the whole symmetric slab (right).

of our implementation is good up to 1024 processes and decreases beyond due to load balancing effects.

4.4.2. Charges fitted for the nanomesh

To reduce complexity, the atoms of the nanomesh have been assigned to different sets. Charges of atoms belonging to the same set are enforced to be identical by constraints. For nm, B and N atoms are divided in four types depending on the height above the topmost Rh layer, see Figure 4.6(a). The first set of atoms defines the wire region, the second and third the rim of the pore and the fourth set the pore itself. The same strategy is used for nm-H shown in Figure 4.6(b). Since the difference in height between wire and pore is only 0.1 Å for the latter, the atoms are simply assigned to pore and wire establishing no further sets for the transition between the two regions.

Table 4.2.: RESP charges fitted for nm and nm-H. Atom types correspond to the assignment in Figure 4.6.

nm		nm-H	
atom type	charge	atom type	charge
B1	0.040	B1	0.014
B2	0.048	B2	0.031
B3	0.074	N1	-0.099
B4	0.201	N2	-0.120
N1	-0.139	Rh1	0.000
N2	-0.154	Rh2	0.048
N3	-0.223	H	-0.019
N4	-0.311		
Rh1	0.005		
Rh2	0.046		
RMS	$1.3{\cdot}10^{-5}$	RMS	$5{\cdot}10^{-6}$
RRMS	$2.4{\cdot}10^{-2}$	RRMS	$8{\cdot}10^{-3}$

As shown in Table 4.2, positive charges are generally obtained for boron and negative charges for nitrogen, which is in agreement with the difference in electronegativity. Comparing the magnitude of charges for B and N atoms, larger values have always been obtained for nitrogen. This is an effect arising from the slightly positive charges of Rh and constraining the total charge to zero. Furthermore, the absolute charge values for B and N are larger in the pore and decrease gradually moving from rim to wire. This is essential to reproduce the modulation of the potential as shown below. The intercalated H atoms of nm-H are negatively charged pointing to a hydride-like character. This seems reasonable because the H monolayer is quasi-bonded to the topmost Rh layer.

Table 4.3.: Difference of the electrostatic potential ΔV (eV) between pore and wire obtained by DFT and RESP charges at different heights above the wire.

	nm		nm-H	
Height [Å]	ΔV_{DFT}	ΔV_{RESP}	ΔV_{DFT}	ΔV_{RESP}
2.8	0.640	0.615	0.156	0.157
3.8	0.535	0.511	0.108	0.119
4.8	0.438	0.411	0.088	0.094

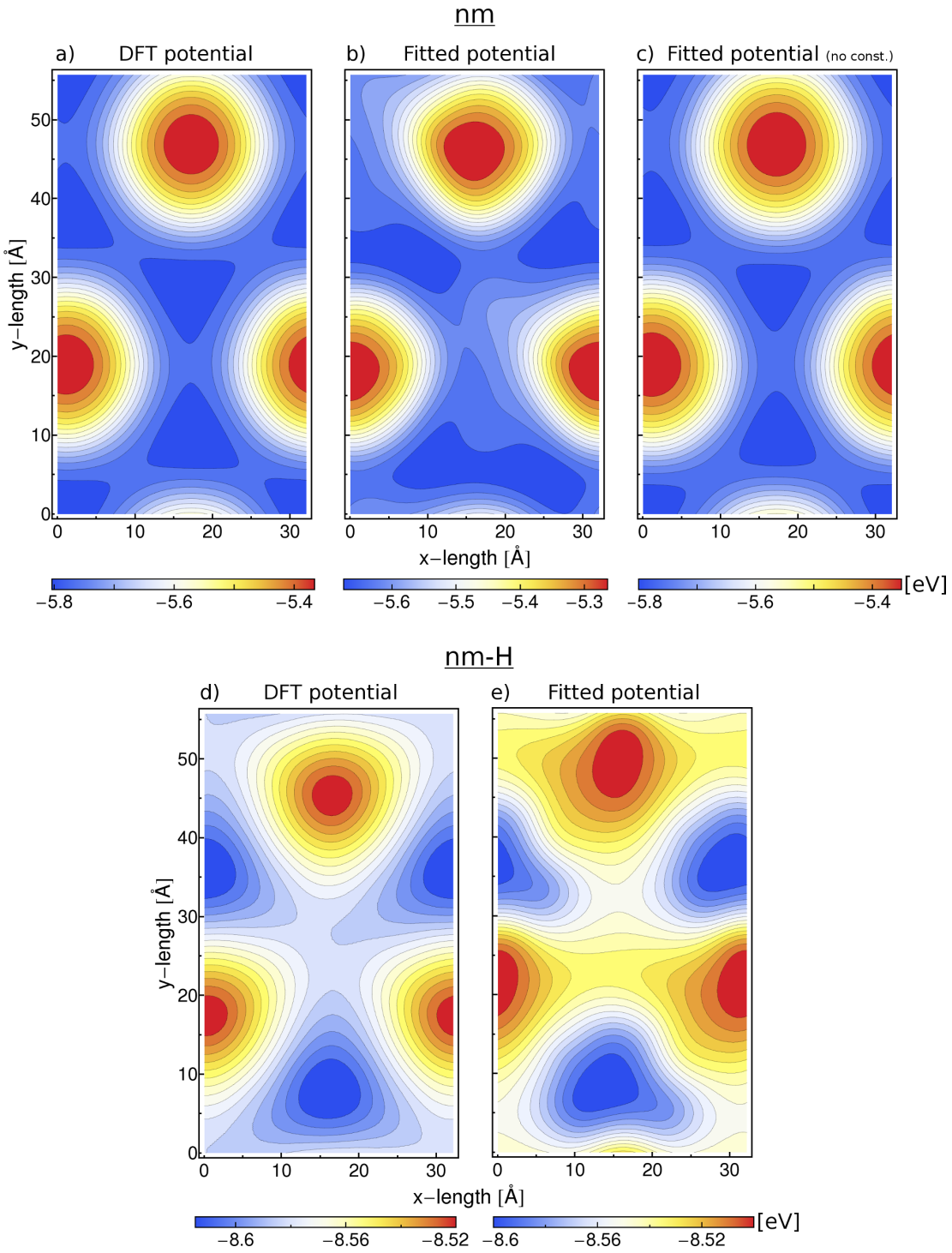

Figure 4.7.: Electrostatic potential at 4.8 Å above the wire. Potential for nm obtained by (a) DFT, (b) generated by RESP charges given in Table 4.2 and (c) potential generated by RESP charges from an unrestrained and unconstrained fit. (d) DFT potential and (e) fitted potential for nm-H.

RMS and RRMS errors are very small indicating a high quality of the fit. The RRMS errors are smaller [189] or at least comparable [190] to the values reported for the RESP fitting of periodic molecular structures. To gain better insight into the characteristics of the ESP, the latter is plotted at a height of 4.8 Å above the wire in Figure 4.7. A strong modulation of the potential is observed, where the absolute values are smaller above the pore than above the wire. The fitted potential reproduces the main features of V_{QM} qualitatively very well. The resemblance seems slightly worse for nm-H. Assigning atoms to different sets and constraining their charges causes slight distortions of the symmetry of the potential, whereas the symmetry is preserved in an unconstrained and unrestrained fitting procedure, see Figure 4.7(c). Since nm-H is almost flat, the assignment of atoms to the pore and wire is somewhat delicate and can strongly affect the modulation and the symmetry of the fitted potential. Quantitatively, the potential difference ΔV between pore and wire is of primary importance and is reported in Table 4.3 for different heights above the wire. ΔV is generally well reproduced by the RESP charges for both systems and it is 4-5 times larger for nm. At 2.8 Å above the surface, the deviation compared to DFT is less than 4% for nm and 1% for nm-H. Further away from the surface - at 4.8 Å -, the deviation is a bit larger, but still less than 7%. Recalling that the potential was solely fitted between 2 to 4 Å above the slab, it is encouraging that even for regions outside the fitting range the potential is well represented by the RESP charges.

4.4.3. Structural and dynamic properties of water at the nm and nm-H interface

The structural properties and the dynamics of the liquid water film interacting with the two different substrates are investigated from the analysis of the MD trajectories. A typical snapshot of the water film on each substrate is presented in Figure 4.9(a and b). Radial distribution functions (RDFs) for different layers of the slab are reported in Figure 4.8 and are essentially identical for both systems. In the middle of the slab, the RDF reproduces the typical shape found for bulk liquid water by DFT-PBE studies [143,149]. The reduced peak heights obtained for RDFs of the interface regions are due to the diminished coordination. The slight difference in peak heights for the latter suggests that water is more structured at the substrate than at the vacuum interface.

The plane-averaged density of the water molecules ρ_{w} as function of the vertical distance from the surface is presented in Figure 4.9(c and d). Since the absolute height of BN changes depending on the corrugation and on the presence of intercalated H, the plane of the highest BN pairs was taken as reference ($z = 0$) to define the distance of the water molecules from the surface. Both systems show a sharp large peak at 3.1 Å (nm) and 3.2 Å (nm-H) where the density is more than twice as large as in bulk water. The adlayer is followed by a 2.0 Å broad region of density depletion with a significantly

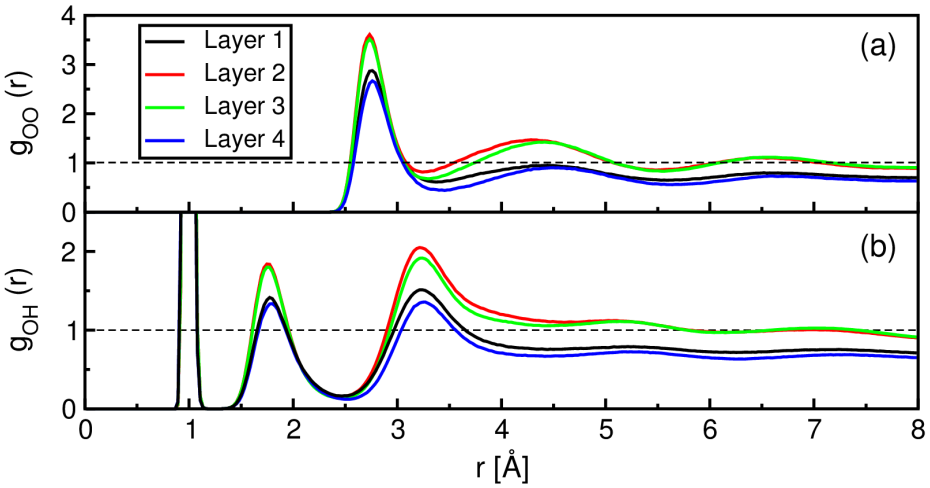

Figure 4.8.: (a) Oxygen-oxygen and (b) oxygen-hydrogen RDFs for water/nm. The four different layers are defined as shown in Figure 4.9(c).

smaller density of 0.5 g/cm^3. A second broad peak is found at 6.0 Å showing that the water configuration is still dictated by the substrate. The first and second peak indicate that the water molecules at the interface are overstructured compared to bulk water. At larger distances, the density distribution is uniform corresponding to bulk water. The density decreases finally to zero at the liquid-vacuum interface. In the case of nm, an earlier onset of the density profile and an additional broad shoulder at 2.3 Å are observed. The 2D plot of the spatial distribution of O atoms at a distance of 2.3 Å, see Figure 4.9(f), shows that water molecules are localized in the pore at this height, while none are found over the wire. This proves that the water film follows the corrugation of the BN layer and confirms the visual impression obtained from the snapshot in Figure 4.9(a).

For further analysis, the water films are partitioned into four layers following the oscillations in the averaged density profile as shown in Figure 4.9(c and d). First information on the hydrogen bonding patterns at the interface are provided by the plane-averaged distributions of O and H atoms presented in Figure 4.9(e). The largest fraction of the hydrogens in the first layer is located in the plane of the O atoms indicated by the large peaks at 3.1 Å (nm) and 3.2 Å (nm-H). At this height, the H atom density is only 1.3 times larger compared to the O atom density, whereas a factor of two is expected for bulk water. This suggests that many H-bonds are formed with the acceptors of the second layer or the substrate. The H-bonds donated to the second layer are counted applying a standard criterion for defining a H-bond [150,151]. In both cases roughly 45% of the water molecules of the adlayer donate one H-bond to the second layer, but only 2% of them donate both. All the other H atoms, which are not involved in a H-bond to a molecule in the second layer, form either an in-layer H-bond, or point towards the substrate. The onset of the H-distribution is for both systems at shorter distances than for oxygen alluding that H atoms point towards

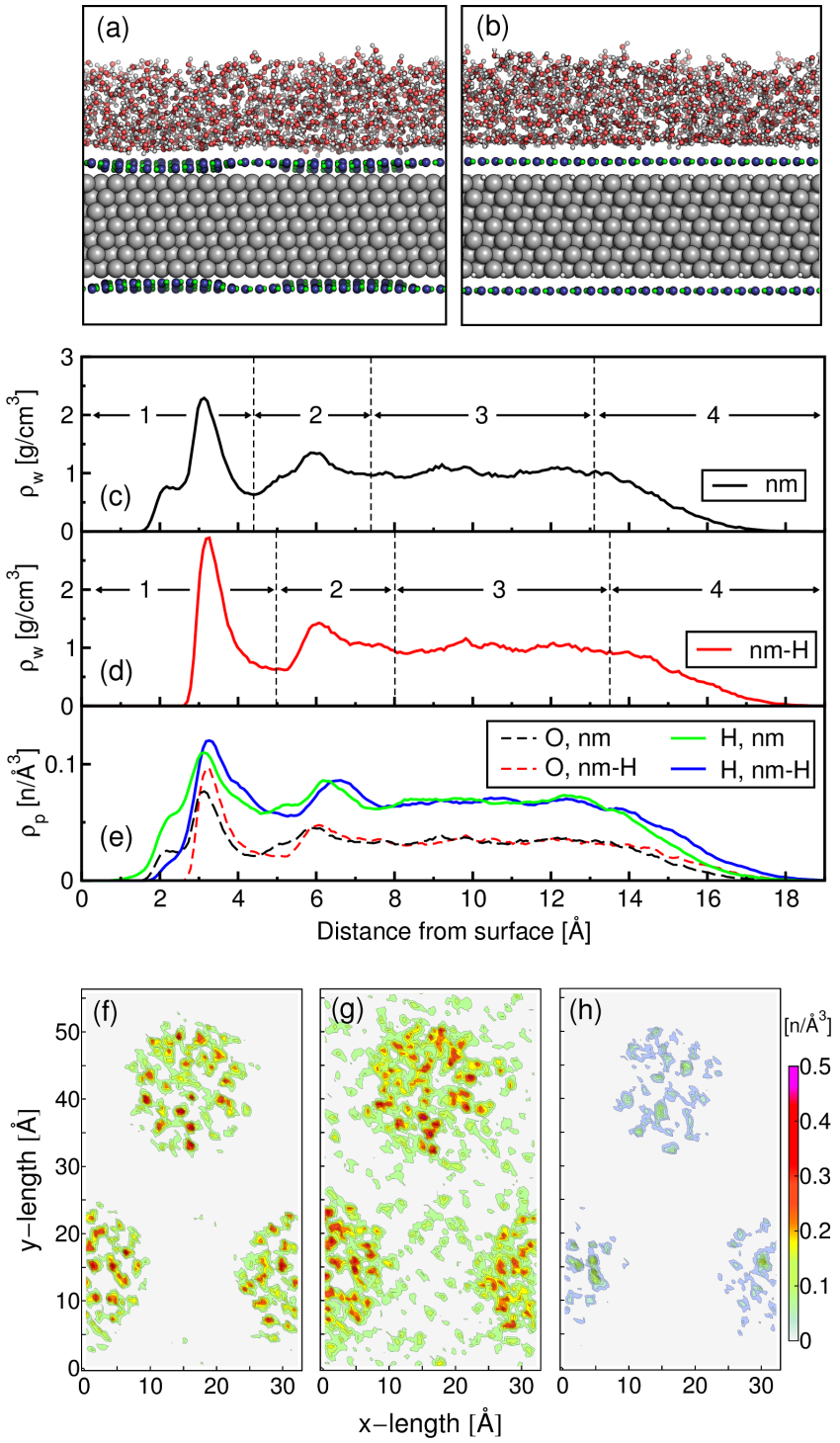

Figure 4.9.: Typical snapshot of the simulation cells of (a) water/nm and (b) water/nm-H. (c-e) Properties as function of the distance from the BN wire. Plane-averaged water density ρ_w along the surface normal (z-direction) for (c) water/nm and (d) water/nm-H. The arrows indicate the division of the water film in four different layers. (e) Plane-averaged particle densities ρ_p for O and H atoms. Spatial distribution of (f) O and (g) H atoms for water/nm at 2.3 Å from the wire. (h) Spatial distribution of H atoms at a distance of 1.2 Å.

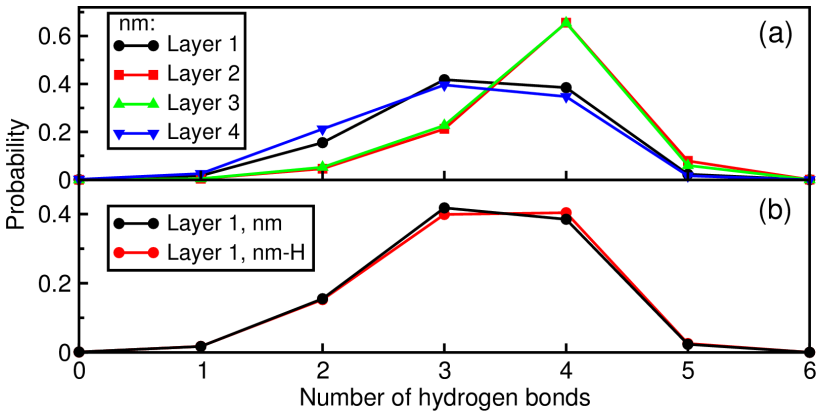

Figure 4.10.: Distribution of the total number of H-bonds per molecule. The layers of the water film are defined in Figure 4.9(c and d). (a) Distribution of H-bonds for water/nm. (b) Distribution of H-bonds at the water-substrate interface comparing water/nm and water/nm-H.

the surface. This is confirmed by the 2D plot of the spatial hydrogen distribution at 1.2 Å displayed in Figure 4.9(h) for water/nm. At distances smaller than 1.5 Å, the probability of finding O atoms is zero, whereas this not true for H atoms. Some water molecules located in the pore direct their hydrogens towards the substrate. This is also observed for molecules above the wire. At 2.3 Å, oxygen can be only found in the pore, but H atoms can be also found above the wire, see Figure 4.9(f and g).

The average number of accepted and donated H-bonds per molecule has been computed for further investigation of the hydrogen bonding properties. The distribution of H-bonds in each layer is plotted for water/nm in Figure 4.10(a). In the bulk-like part of the film, the second and third layer, most of the molecules are involved in four hydrogen bonds, which is in agreement with DFT-PBE studies of bulk water [143,146,203,204]. At the interface, the H-bonding is frustrated and the probability for forming three H-bonds is slightly larger than for a tetrahedral configuration. Small differences between the H-bond distributions of the adlayer are observed comparing the two substrates. As shown in panel (b) of Figure 4.10, the probabilities for three and four H-bonds are equivalent for the water/nm-H system indicating slightly less frustrated H-bond patterns imposed by the substrate.

Indication of a possible polarization of the molecules interacting with the substrate could be obtained by the analysis of the average orientation of the dipole moment. As first approximation to the dipole moment orientation, we consider the angle α between surface normal and the vector $\mathbf{d}$ lying in the H_2O molecular plane and bisecting the HOH angle, as shown in Figure 4.11(a). An angle of 90° refers to a dipole orientation parallel to the surface. The vector $\mathbf{d}$ points away from the substrate if $0° \leq \alpha < 90°$ and towards the substrate if $90° < \alpha \leq 180°$. The distribution of α is shown for water/nm in Figure 4.11(b). Note that a homogeneous orientational distribution exhibits the shape of a sine with a maximum at 90°, which can be validated by simple

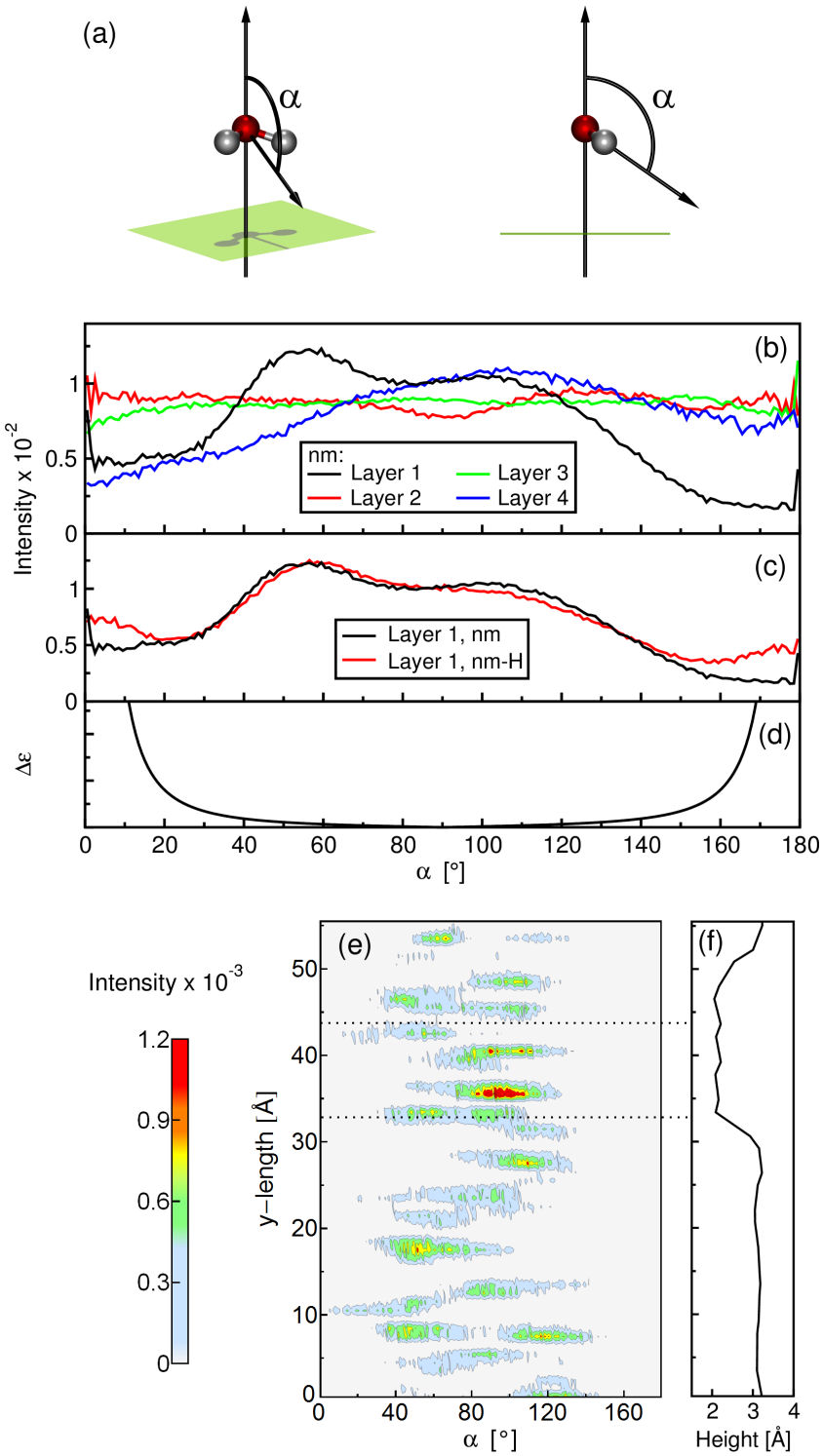

Figure 4.11.: Orientation of the water molecules. (a) Definition of the orientation angle α with respect to the surface normal. (b) Distribution of α for each layer obtained for water/nm. (c) Distribution of α in the adlayer for water/nm and water/nm-H. (d) Absolute error $\Delta\epsilon$ for the distribution of α. (e) Angle α correlated with the spatial distribution of water molecules along the horizontal y-axis of the orthorhombic unit cell of nm. (f) Height of the BN layer measured relative to the topmost Rh layer along the y-axis.

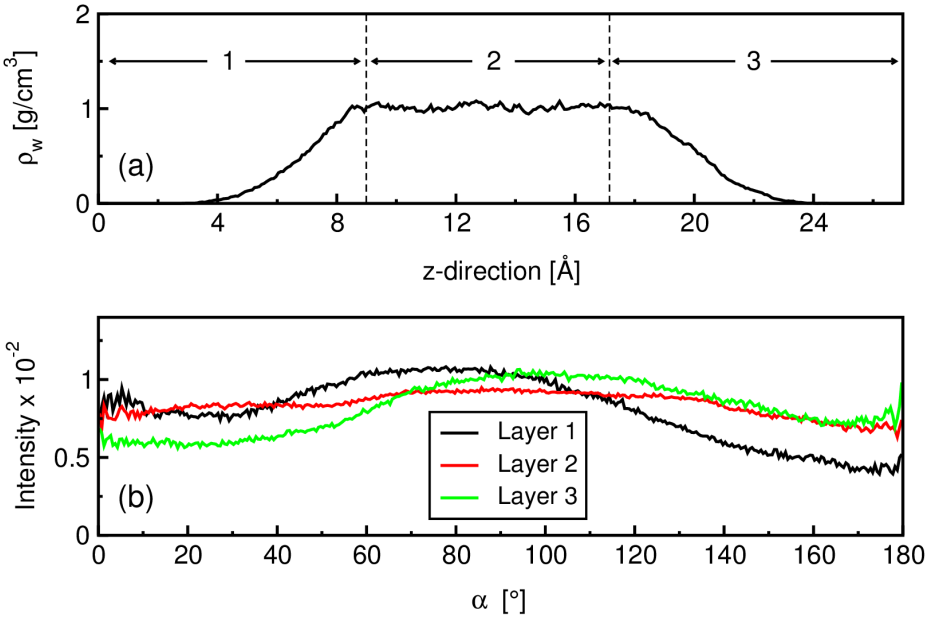

Figure 4.12.: Results for the pure water slab. (a) Plane-averaged water density ρ_w along the z-direction which is perpendicular to the water-vacuum interface. The arrows indicate the division of the water film in three different layers. (b) Distribution of the orientation angle α for the three different layers.

trigonometric considerations. In this representation, it is difficult to evaluate if any angle is preferred. Therefore, the distributions have been normalized by $\sin\alpha$ such that for homogeneous distributions no or only a weak dependence on α is expected. Homogeneous distributions are found for layers two and three confirming their bulk-like nature, see Figure 4.11(b). At the liquid/vacuum interface, the distribution has a broad peak at around 105° indicating that the dipole moments are preferentially directed towards the bulk. The same is observed for the liquid/vacuum interfaces of a pure water slab, see Figure 4.12. This is understood with the tendency of water to optimize the number of H-bonds with other water molecules. Two peaks are instead observed for molecules belonging to the adlayer: one at 55° indicating an increased probability for the orientation towards the bulk and a second one at around 100°−110°, which corresponds to molecules turned towards the substrate. Molecule orientations towards the surface are associated with increased interactions between adlayer and substrate mediated by H-bonds.

Comparing the distributions in the first layer for water/nm and water/nm-H, we find that the orientation angle of $100° - 110°$ is less favored for water/nm-H, see Figure 4.11(c). This implies that the interactions with the surface are weaker for the flat nanomesh. The curves displayed in Figure 4.11(c) deviate also for $\alpha < 30°$ and $\alpha > 150°$. These differences must be attributed to technical artifacts. Due to limited sampling and the discrete nature of the distribution, each angle is error-prone. The errors are especially large for small and large angles, which is illustrated by $\Delta\epsilon$

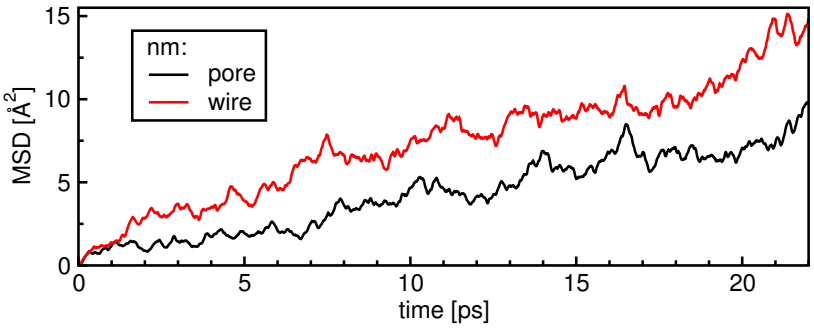

Figure 4.13.: MSD for water/nm observed for a set of 23 molecules of the adlayer located in the pore and over the wire, respectively.

depicted in Figure 4.11(d). The functional form of the absolute error $\Delta\epsilon(\alpha)$ is given by $\left|\frac{\partial(\sin\alpha)^{-1}}{\partial\alpha}\right|$.

We also want to determine whether the molecules orientated towards the substrate are preferentially located at specific sites of the nanomesh, i.e. pore or wire. For that reason, α is correlated with the spatial distribution of water molecules along the y-dimension of the unit cell at a fixed x-value, see Figure 4.11(e). In Figure 4.11(f), the corresponding h-BN height along the same line is reported, thus identifying the pore area between $30-50$ Å. We find that angles of $95°-110°$ are more probable for water molecules in the pore pointing to stronger interactions here. This agrees well with previous DFT based studies where stronger interactions between water molecules and nanomesh are predicted for molecules trapped in the pore [182].

In order to estimate the diffusivity of the water molecules on nm, the mean square displacement [78] (MSD) has been computed and is presented in Figure 4.13. The MSD has been measured for two sets of 23 water molecules belonging to the first layer over the pore and over the wire, respectively. The slope of the time-dependent MSD plot for molecules on the wire is clearly larger than for molecules in the pore. This indicates that the diffusivity within the pore is substantially reduced confirming that the modulated ESP tends to trap molecules there. These results agree with experimental and theoretical studies of nano-ice clusters on nm [180, 184]. Based on scanning tunneling microscopy, Ma *et al.* [180] proposed the formation of stable ice hexamers in the pore, whereas well-defined structures have not been found on the wire. The latter has been interpreted as result of the high mobility of weakly adsorbed molecules.

4.4.4. Wetting behavior

To assess the wetting behavior of water on nm and nm-H, interaction energies E_{int} and properties related to the contact angle θ have been computed, see Table 4.4. The interaction energy is 4% larger for water/nm indicating that the surface is more hy-

Table 4.4.: Interaction energy E_{int} (kJ/(mol·nm^2)) and ratio of solid-liquid γ_{SL} with respect to liquid-vacuum energy γ_{LV}.

	E_{int}	$-\frac{\gamma_{SL}}{\gamma_{LV}}$
water/nm	−90.9	1.026
water/nm-H	−87.4	0.987
Δ	−3.5	0.04

drophilic. Macroscopically, smaller contact angles correlate with a larger hydrophilicity of the surface. The contact angle θ is not directly accessible, but it can be related to the ratio $-\gamma_{SL}/\gamma_{LV}$ defined in Section 4.3.2. A larger ratio corresponds to a larger $\cos\theta$ and a smaller θ, given that $0° \leq \theta \leq 180°$. From our evaluation of the liquid/substrate interaction energies, we obtain that the ratio $-\gamma_{SL}/\gamma_{LV}$ is larger for water/nm predicting consequently a smaller contact angle. Taking the difference of the ratios, the resulting $\Delta\cos\theta$ is 0.04, which is a property that could potentially be confirmed by experimental studies. However, the comparison to experiment is not straightforward due to the approximations implied in our model. First, the Young equation is an approximation assuming a static contact angle for a droplet on a surface, whereas in practice a range of angles is observed [200]. Also it is assumed that γ_{LV} is the same for the two systems. Second, our theoretical model for describing the water/substrate interface is limited since the interactions between liquid and surface are treated at the MM level of theory. The electrostatic interactions are modeled accurately as demonstrated in Section 4.4.2. However, the dispersion interactions might be slightly overestimated which has been discussed in Section 4.4.1.

4.5. Conclusions

We introduced a method to fit RESP charges for periodic systems based on the GPW scheme and we applied it to study the wetting behavior of water on the *h*-BN/Rh nanomesh depending on the modulation of the ESP. The proposed periodic RESP method offers great flexibility in assigning the real space volumes where the reference ESP should be reproduced with higher accuracy. This allows to tune the calculation of RESP charges easily for nano-structured materials exposing surfaces and porous structures.

RESP charges have been determined to reproduce the ESP of two different substrates based on the *h*-BN/Rh interface, i.e. nm and nm-H. The first is characterized by a relative strong modulation of the potential, following the corrugation of the *h*-BN layer, while the effect for the second is reduced due to the presence of intercalated

H atoms. The resulting charges have been used to model the interaction of the two substrates with liquid water within a QM/MM approach, where water is treated at the DFT level and the interaction with the substrate at the MM level.

Structural and dynamic properties of the two interfaces have been extracted from the analysis of the generated MD trajectories. We observe that the water film follows the corrugation of the *h*-BN layer for nm. In both cases the interaction of water with the substrate is responsible for the formation of an overstructured adlayer with an unfavorable H-bond network with respect to water's preferred tetrahedral structure. The effect of the frustration of the H-bonding appears to be slightly more pronounced at the nm than at the nm-H interface. By inspecting the average orientation of water molecules, an increased probability is found for molecules turned towards the substrate, which is in particular associated to molecules placed over the pore of nm. In addition to that, molecules close to the pore show a reduced diffusivity with respect to molecules staying over the wire.

Finally, interaction energies between substrate and water film and the surface energy of a free standing liquid water film have been calculated. These energies are used to compare the wetting of water on nm and nm-H. We find that the nm surface is slightly more hydrophilic, as also shown by the structural investigation. Macroscopically, we predict a smaller contact angle for a droplet of water on nm than on nm-H.

5. Local density fitting within a Gaussian and plane waves approach

A local resolution of identity approach (LRI) is introduced within the Gaussian and plane waves (GPW) framework. The construction of the Kohn-Sham (KS) matrix in GPW scales already linearly with respect to system size by using a plane wave expansion of the density for the evaluation of the Coulomb term in combination with a local basis. The intention is to retain the linear scaling of the GPW approach, while reducing the prefactor for computing the KS matrix. The locality of the density fitting ensures an $O(n)$ scaling and is implemented by approximating the atom pair density by an expansion in one-center fit functions. The prefactor is smaller with LRI since the computational demands for the grid-based operations become negligible, while they are dominant in GPW. The accuracy of LRIGPW is assessed for many different systems and properties. Generally, total energies, reaction energies, intramolecular and intermolecular structure parameters are well reproduced. LRIGPW yields also high quality results for extended periodic structures such as liquid water, ice XV and molecular crystals. The speed-up of the self-consistent field (SCF) procedure depends mainly on the symmetry of the simulation cell, the grid cutoff and the system size. A speed-up of the SCF step by a factor of 2-25 has been observed for periodic systems.

5.1. Introduction

The computational cost of density functional theory (DFT) calculations is mainly dominated by the evaluation of the Coulomb term J. The Coulomb term requires the calculation of the four-center two-electron repulsion integrals (ERIs), defined in Equation 2.29. The computational cost for the calculation of the ERIs grows with the fourth power of the number of basis functions n. The $O(n^4)$ scaling behavior becomes a bottleneck for large systems and will limit the system size.

Density fitting approaches, also referred to as resolution of identity (RI), were introduced in order to accelerate the evaluation of J. Early attempts aimed at approximating the two-center product $\chi_\mu^{\mathrm{A}}\chi_\nu^{\mathrm{B}}$, where χ_μ^{A} and χ_ν^{B} are basis functions centered at A and B, respectively [61–66, 205]. Pioneering work on approximating the full density [67, 68, 206] (global fitting) or the atomic pair densities [207] (local fitting) by an expansion in one-center auxiliary fit functions dates back to the 1970s. Global fitting schemes were first proposed by Sambe and Felton [67] and further developed

by Dunlap [68], Ahlrichs et *al.* [45,70,71] and others [69]. Using global RI, the scaling of the ERIs is formally only reduced to $O(n^3)$, since the fitting process itself is an $O(n^3)$ process. However, the prefactor is significantly smaller leading to speed-ups of more than a factor of 10 [71].

In local fitting schemes on the contrary, the fitting of the expansion coefficients scales formally only quadratically and can be further reduced to $O(n)$. Thus, linear-scaling schemes for J with respect to molecular size can be obtained. Local fitting has been also employed for the exact exchange contribution K in hybrid functionals [75,208,209]. Depending on the algorithm, the scaling of the K term is not necessarily reduced, but merely the prefactor [209]. Locality in density fitting is achieved by either using local metrics [75, 207] or local domains [50, 208, 210]. The latter rely on pre-defined criteria to localize the fitting region. For example, Sodt *et al.* [208] proposed a scheme where each product of orbital functions is tagged to a parent atom and fitted using only auxiliary functions on atoms in a certain radius around this parent atom. Local metrics define an alternative local operator. Baerends *et al.* [207] introduced a local metric where the electron density is decomposed into pair-atomic densities ρ_{AB}. Each of these pair densities are approximated by an expansion in auxiliary functions centered at atoms A and B. In the early work of Baerends and Ros, an overlap metric is employed, i.e. the residual of the overlap is minimized. This strategy was subsequently used by several others [46,211] and also in this work.

An overlap metric yields linear errors, whereas the error in the fit is quadratic for Coulomb metrics where the Coulomb repulsion of the density residual is minimized [68,75]. Several attempts were made to introduce locality for the latter. Attenuated Coulomb metrics have been developed by Jung *et al.* [212] and Reine *et al.* [213] using damping parameters that lead to rapid decays of the expansion coefficients with interelectronic distance. Recently, Merlot *et al.* [75] proposed the Pair Atomic Resolution of identity (PARI) approximation. In this approach, the density is split in a sum of atom pair densities as suggested by Baerends *et al.* [207]. The major difference is that each product $\chi_\mu^{\mathrm{A}}\chi_\nu^{\mathrm{B}}$ for each pair AB is fitted instead of the pair densities ρ_{AB}. Furthermore, the Dunlap correction [214] and a Coulomb metric is employed. The advantage compared to Baerends' approach is that the fitting is robust, variational and the error of the fit is quadratic. However, the prefactor for the fitting is larger since each product $\chi_\mu^{\mathrm{A}}\chi_\nu^{\mathrm{B}}$ for each pair AB is approximated. A further obstacle is that the PARI approximation for the J term leads to unphysical attractive electronic interactions and gives rise to severe convergence problems.

Apart from an $O(n)$ scaling, the local atomic-pair fitting approach proposed by Baerends *et al.* has several advantages compared to global RI. The quality of the fitted pair density does not depend on the chemical environment, because the fitting is always restricted to pair AB. Since the fitting is applied separately to each pair, the dimensions remain in principle modest and are independent of the system size. Another advantage is that pair-wise fitting schemes can be easily parallelized by dis-

tributing the pairs over the nodes. However, since the error in the fit is linear, larger auxiliary basis sets have to be used to reach the same accuracy as with Coulomb metrics. Fitting each pair AB separately, the presence of auxiliary functions at neighbor C cannot be exploited increasing additionally the size of the auxiliary basis sets compared to global fitting techniques.

In this work, the local resolution of identity (LRI) approach of Baerends *et al.* [207] has been adapted to a Gaussian and plane waves (GPW) framework [44, 118], implemented in the CP2K program package [121,191] and is in the following referred to as LRIGPW. In CP2K, the evaluation of the Coulomb interaction and the construction of the Kohn-Sham (KS) matrix scale already linearly with system size by exploiting the plane wave expansion of the density to solve the Poisson equation in Fourier space. In combination with a local basis, it is possible to reduce the scaling for the computation of the Coulomb matrix to an $O(n)$ process. Using LRI, the scaling is still linear, while the prefactor for the calculation of the grid-dependent terms is massively reduced. The fitted density is computed on the grid and subsequently used for evaluation of the core-electron interactions and Coulomb and exchange-correlation contributions.

In the next section, the theory and implementation details of LRIGPW are described. The construction of the auxiliary basis sets and their optimization are discussed. The accuracy of LRIGPW is assessed by comparing total and reaction energies as well as molecular structure parameters for a broad range of systems. LRIGPW is employed to evaluate different properties of periodic structures such as ice XV and molecular crystals. Molecular dynamics (MD) simulations of liquid water have been performed as well. Finally, timings and remaining problems are discussed.

5.2. The LRIGPW method

5.2.1. Central aspects and bottlenecks of the GPW method

In GPW, a dual representation of the electronic density $\rho(\mathbf{r})$ is used [44, 118]. The density is expressed in terms of a local basis of atom-centered, contracted Gaussian functions $\{\chi_\mu\}$,

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}), \tag{5.1}$$

where $P_{\mu\nu}$ is the density matrix. Alternatively, the density can be represented in an auxiliary basis set of plane waves (PW),

$$\rho(\mathbf{r}) = \frac{1}{\Omega}\sum_{\mathbf{G}} \rho(\mathbf{G})\exp(i\mathbf{G}\cdot\mathbf{r}), \tag{5.2}$$

where Ω is the volume of the unit cell, $\mathbf{G}$ the reciprocal lattice vector and $\rho(\mathbf{G})$ denotes the representation of the density in reciprocal space. The density is mapped

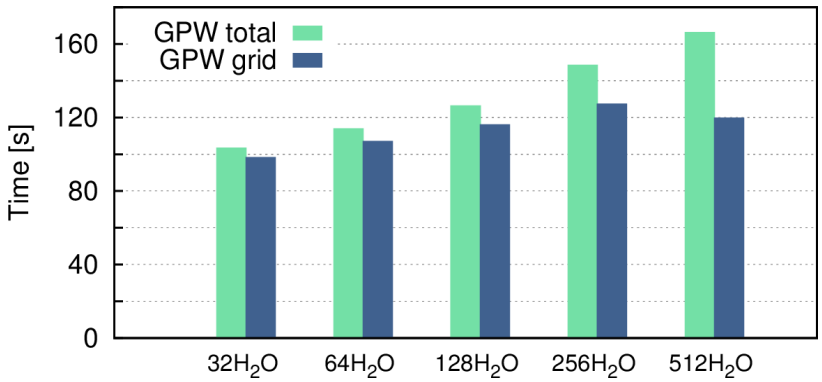

Figure 5.1.: Total GPW run time and time needed for the operations on the grid (collocation of the density and integration of the potential). Calculations are carried out for liquid water boxes of different sizes with the setup described in Section 5.4.

on a regular grid and Fast Fourier Transforms (FFTs) are used to rapidly convert the representation from real to reciprocal space and vice versa.

The Coulomb energy $J[\rho]$ in the PW representation is given by

$$J[\rho] = 2\pi\Omega \sum_{\mathbf{G}} \frac{\rho^*(\mathbf{G})\rho(\mathbf{G})}{\mathbf{G}^2} = \frac{\Omega}{2} \sum_{\mathbf{G}} \rho^*(\mathbf{G}) V_{\mathrm{H}}(\mathbf{G}), \tag{5.3}$$

where $V_{\mathrm{H}}(\mathbf{G}) = 4\pi\rho(\mathbf{G})/\mathbf{G}^2$ is the Hartree potential in reciprocal space. $V_{\mathrm{H}}(\mathbf{G})$ is obtained by collocating the density on a real space grid and transferring it to reciprocal space by FFT. The Poisson equation is solved in reciprocal space and the potential is subsequently converted to its real space analogue $V_{\mathrm{H}}(\mathbf{r})$. The XC energy is calculated re-using the real space density grid. The XC potential $V_{\mathrm{xc}}(\mathbf{r})$ is evaluated at each grid point and added to $V_{\mathrm{H}}(\mathbf{r})$. The corresponding contribution $H_{\mu\nu}$ to the KS matrix is obtained by integrating the sum of both potentials on the grid,

$$H_{\mu\nu} = \int \left(V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r})\right) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}. \tag{5.4}$$

Within the GPW framework, the calculation of the potential and of the KS matrix scales linearly with system size. This is achieved in two steps. Namely, using a local basis of Gaussian functions, the expression of the KS density introduced in Equation 5.1 can be regarded as a sum of atomic pair densities $\rho_{\mathrm{AB}}(\mathbf{r})$,

$$\rho(\mathbf{r}) = \sum_{\mathrm{AB}} \rho_{\mathrm{AB}}(\mathbf{r}) = \sum_{\mathrm{AB}} \sum_{\mu\in\mathrm{A},\nu\in\mathrm{B}} P_{\mu\nu} \chi_\mu^{\mathrm{A}}(\mathbf{r}) \chi_\nu^{\mathrm{B}}(\mathbf{r}). \tag{5.5}$$

The first step is to discard all negligible contributions to $\rho(\mathbf{r})$ by exploiting effective screening procedures. Only non-zero products of Gaussian functions on atoms that are sufficiently close are included, i.e. only pairs AB, where B is a neighbor of A, are considered. The number of these neighbor pairs scales linearly with system size. Second, the scaling of $J[\rho]$ is reduced to $O(n)$ by employing a PW representation

and solving the Poisson equation in reciprocal space as described above. However, the prefactor of the grid-based operations, i.e. the collocation of the density and the integration of the potential on the grid, is rather large for periodic systems. The reason is that these grid-based operations have to be performed for each pair of atoms AB stored in the neighbor list. For a liquid water box of 64 H_2O, the number of pairs is already larger than 200 000. The relation between total run time and grid-based operations is displayed in Figure 5.1 for water boxes of different sizes. For the medium-sized systems up to 128 H_2O molecules, the grid-operations require more than 90% of the total run time. Even for larger systems, where matrix operations become more time-consuming, the grid operations are the most expensive steps.

The objective of this work is to reduce the prefactor for the grid-based operations by introducing a density fitting scheme. The fitting scheme needs to be local to retain the linear scaling for the KS matrix. An overlap metric is the most reasonable choice since the fitted density is employed for both, Coulomb and XC part. A Coulomb metric is superior in terms of accuracy, but is applied to the Coulomb part only. These requirements, locality in combination with an overlap metric, are fulfilled by the pair-fitting scheme proposed by Baerends *et al.* [207]. This scheme is opportunely adapted to the GPW scheme and from now on named the LRIGPW method.

5.2.2. Local density fitting

Within the LRIGPW approach, the atomic pair densities ρ_{AB} are approximated by an expansion in a set of fit functions centered at atom A $\{f_i^{\mathrm{A}}\}$ and atom B $\{f_j^{\mathrm{B}}\}$,

$$\tilde{\rho}_{\mathrm{AB}}(\mathbf{r}) = \sum_i a_i^{\mathrm{A}} f_i^{\mathrm{A}}(\mathbf{r}) + \sum_j a_j^{\mathrm{B}} f_j^{\mathrm{B}}(\mathbf{r}). \tag{5.6}$$

The fit functions are also Gaussian-type functions and provided as an additional (auxiliary) basis set. The expansion coefficients $\left\{a_i^{\mathrm{A,(AB)}}\right\}$ and $\left\{a_j^{\mathrm{B,(AB)}}\right\}$ are obtained for each pair AB. Note that we drop the upper index (AB) except for ambiguous cases. The expansion coefficients are obtained by minimizing the difference D between exact and approximated density $\tilde{\rho}_{\mathrm{AB}}$,

$$D_{\mathrm{AB}} = \int |\rho_{\mathrm{AB}} - \tilde{\rho}_{\mathrm{AB}}|^2 d\mathbf{r}, \tag{5.7}$$

imposing the restrictive condition that the integral of the pair density, N_{AB}, remains constant, i.e.,

$$N_{\mathrm{AB}} = \int \tilde{\rho}_{\mathrm{AB}}(\mathbf{r}) d\mathbf{r} = \int \rho_{\mathrm{AB}}(\mathbf{r}) d\mathbf{r} = \sum_{\mu\in\mathrm{A},\nu\in\mathrm{B}} P_{\mu\nu} \int \chi_\mu^{\mathrm{A}}(\mathbf{r}) \chi_\nu^{\mathrm{B}}(\mathbf{r}) d\mathbf{r}, \tag{5.8}$$

where $\{\chi_\mu\}$ denote functions of the primary basis set. This corresponds to a constraint on the number of electrons N_{elec} since it holds that $N_{\mathrm{elec}}{=}\sum_{\mathrm{AB}} N_{\mathrm{AB}}$. Minimizing Equation 5.7 with respect to the expansion coefficients and applying the charge constraint,

yields a linear set of equations for each pair,

$$\mathbf{Sa} = \mathbf{t} + \lambda \mathbf{n}, \tag{5.9}$$

where $\mathbf{a}$ is the vector of the expansion coefficients and $\mathbf{S}$ is the overlap of the auxiliary basis functions,

$$S_{ij} = \int f_i^{\mathrm{A}} f_j^{\mathrm{B}} d\mathbf{r}. \tag{5.10}$$

The elements on the right hand side (rhs) of Equation 5.9 are

$$t_i = \sum_{\mu \in \mathrm{A}, \nu \in \mathrm{B}} P_{\mu\nu} T_{\mu\nu i} \tag{5.11}$$

with

$$T_{\mu\nu i} = \int \chi_\mu^{\mathrm{A}} \chi_\nu^{\mathrm{B}} f_i^{\mathrm{A/B}} d\mathbf{r}, \tag{5.12}$$

where the product $\chi_\mu^{\mathrm{A}} \chi_\nu^{\mathrm{B}} f_i^{\mathrm{A}}$ is integrated for fit functions centered at A and $\chi_\mu^{\mathrm{A}} \chi_\nu^{\mathrm{B}} f_i^{\mathrm{B}}$ for functions at B. The last term on the rhs of Equation 5.9 is associated with the constraint, where the vector n is given by

$$n_i = \int f_i^{\mathrm{A/B}} d\mathbf{r} \tag{5.13}$$

and is non-zero only for s-functions. The Lagrange multiplier λ is given by

$$\lambda = \frac{N_{\mathrm{AB}} - \mathbf{n}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{t}}{\mathbf{n}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{n}}. \tag{5.14}$$

The complete derivation of Equation 5.9 can be found in Appendix B.1. In matrix representation, the linear system of equations is given by

$$\begin{pmatrix} S_{11}^{\mathrm{AA}} & \cdots & S_{1m}^{\mathrm{AA}} & S_{11}^{\mathrm{AB}} & \cdots & S_{1k}^{\mathrm{AB}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{m1}^{\mathrm{AA}} & \cdots & S_{mm}^{\mathrm{AA}} & S_{m1}^{\mathrm{AB}} & \cdots & S_{mk}^{\mathrm{AB}} \\ S_{11}^{\mathrm{BA}} & \cdots & S_{1m}^{\mathrm{BA}} & S_{11}^{\mathrm{BB}} & \cdots & S_{1k}^{\mathrm{BB}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{k1}^{\mathrm{BA}} & \cdots & S_{km}^{\mathrm{BA}} & S_{k1}^{\mathrm{BB}} & \cdots & S_{kk}^{\mathrm{BB}} \end{pmatrix} \begin{pmatrix} a_1^{\mathrm{A}} \\ \vdots \\ a_m^{\mathrm{A}} \\ a_1^{\mathrm{B}} \\ \vdots \\ a_k^{\mathrm{B}} \end{pmatrix} = \begin{pmatrix} t_1^{\mathrm{A}} \\ \vdots \\ t_m^{\mathrm{A}} \\ t_1^{\mathrm{B}} \\ \vdots \\ t_k^{\mathrm{B}} \end{pmatrix} + \lambda \begin{pmatrix} n_1^{\mathrm{A}} \\ \vdots \\ n_m^{\mathrm{A}} \\ n_1^{\mathrm{B}} \\ \vdots \\ n_k^{\mathrm{B}} \end{pmatrix} \tag{5.15}$$

where m and k are the number of auxiliary basis functions at A and B, respectively. The upper part of the vector $\mathbf{a}$ yields the expansion coefficients on center A and the lower part on center B. The overlap matrix $\mathbf{S}$ includes next to the elements S_{ij}^{AB} and also the self-overlap S_{il}^{AA} and S_{pj}^{BB}. For the self-pair AA, only the upper part of the vectors and the self-overlap S_{il}^{AA} remain.

Based on Equation 5.5, the total fitted density $\tilde{\rho}(\mathbf{r})$ is defined as

$$\tilde{\rho}(\mathbf{r}) = \sum_{\mathrm{AB}} \tilde{\rho}_{\mathrm{AB}}(\mathbf{r}) = \sum_{\mathrm{AB}} \left[\sum_i a_i^{\mathrm{A},(\mathrm{AB})} f_i^{\mathrm{A}}(\mathbf{r}) + \sum_j a_j^{\mathrm{B},(\mathrm{AB})} f_j^{\mathrm{B}}(\mathbf{r}) \right]. \tag{5.16}$$

The expansion coefficients can now be summed up for each atom X. For atom A we have,

$$\tilde{a}_i^{\mathrm{A}} = \sum_{\mathrm{B}\in\mathrm{AB}} a_i^{\mathrm{A},(\mathrm{AB})}, \tag{5.17}$$

where the sum is running over the number of neighbors of A, i.e., we get contributions from the pairs AB, AC, AD etc. Using Equation 5.17, the total density is given by

$$\tilde{\rho}(\mathbf{r}) = \sum_{\mathrm{X}} \sum_{i} \tilde{a}_i^{\mathrm{X}} f_i^{\mathrm{X}}(\mathbf{r}). \tag{5.18}$$

In LRIGPW, the total density is thereby presented as a sum over the number of atoms, whereas in GPW we have a sum over pairs, see Equation 5.5. Coming back to the $64H_2O$ box, the collocation of the fitted density is obtained by performing a loop over 192 atoms rather than 200 000 pairs.

In the fitting scheme proposed here, no constraint is set that prevents the fitted density from getting negative. Indeed, numerical inaccuracies of the fitting procedure can produce small negative values of the density at grid points far from the atomic centers, where the density should be zero. However, the absolute values of these numerical errors are smaller than 10^{-6} a.u., which is less than the error introduced by the interpolation/extrapolation operations associated with the multigrid approach in CP2K [44].

5.2.3. Construction of the Kohn-Sham matrix

The expression for the density-dependent terms of the KS matrix $H_{\mu\nu}$, see Equation 5.4, has to be substituted by its LRIGPW analogue $\tilde{H}_{\mu\nu}$,

$$\tilde{H}_{\mu\nu} = \frac{dE(\tilde{\rho})}{dP_{\mu\nu}} = \sum_i \int \frac{dE(\tilde{\rho})}{d\tilde{\rho}} \frac{d\tilde{\rho}}{da_i^{\mathrm{A}}} \frac{da_i^{\mathrm{A}}}{dP_{\mu\nu}} d\mathbf{r} + \sum_j \int \frac{dE(\tilde{\rho})}{d\tilde{\rho}} \frac{d\tilde{\rho}}{da_j^{\mathrm{B}}} \frac{da_j^{\mathrm{B}}}{dP_{\mu\nu}} d\mathbf{r} \tag{5.19}$$

where $\mu \in \mathrm{A}$ and $\nu \in \mathrm{B}$. Since $dE(\tilde{\rho})/d\tilde{\rho} = \tilde{V}_{\mathrm{H}}(\mathbf{r}) + \tilde{V}_{\mathrm{xc}}(\mathbf{r}) = \tilde{V}(\mathbf{r})$, the expression is simplified to

$$\tilde{H}_{\mu\nu} = \sum_i \frac{da_i^{\mathrm{A}}}{dP_{\mu\nu}} I_i^{\mathrm{A}} + \sum_j \frac{da_j^{\mathrm{B}}}{dP_{\mu\nu}} I_j^{\mathrm{B}}, \tag{5.20}$$

where I_i^{X} is the integral of the potential multiplied by the auxiliary function f_i on atom X,

$$I_i^{\mathrm{X}} = \int \tilde{V}(\mathbf{r}) f_i^{\mathrm{X}}(\mathbf{r}) d\mathbf{r}. \tag{5.21}$$

The derivatives of the expansion coefficients with respect to the density matrix are

$$\begin{aligned} \frac{da_i^{\mathrm{A/B},(\mathrm{AB})}}{dP_{\mu\nu}} &= \frac{d}{dP_{\mu\nu}} \left[\sum_j S_{ij}^{-1} t_j + \lambda \sum_j S_{ij}^{-1} n_j \right] \\ &= \sum_j S_{ij}^{-1} T_{\mu\nu j} + \frac{d\lambda}{dP_{\mu\nu}} \sum_j S_{ij}^{-1} n_j \end{aligned} \tag{5.22}$$

with

$$\frac{d\lambda}{dP_{\mu\nu}} = \frac{\int \chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})d\mathbf{r}}{\mathbf{n}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{n}} - \sum_j \left(\frac{\mathbf{n}^{\mathrm{T}}\mathbf{S}^{-1}}{\mathbf{n}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{n}}\right)_j T_{\mu\nu j}. \tag{5.23}$$

The integral of the potential I_i^{X} is computed for each atom X, whereas in GPW the potential has to be integrated for each pair.

5.2.4. Forces and stress

For the LRIGPW scheme, also the expressions for the forces have to be adapted. The derivatives with respect to $\mathbf{R}_{\mathrm{A}}$ are

$$\begin{aligned}\frac{dE(\tilde{\rho})}{d\mathbf{R}_{\mathrm{A}}} = \sum_i \int \frac{\partial E}{\partial \tilde{\rho}} \frac{\partial \tilde{\rho}}{\partial f_i^{\mathrm{A}}} \frac{\partial f_i^{\mathrm{A}}}{\partial \mathbf{R}_{\mathrm{A}}} d\mathbf{r} + \sum_{\mathrm{B}\in\mathrm{AB}} \sum_i \int \frac{\partial E}{\partial \tilde{\rho}} \frac{\partial \tilde{\rho}}{\partial a_i^{\mathrm{A}}} \frac{\partial a_i^{\mathrm{A}}}{\partial \mathbf{R}_{\mathrm{A}}} d\mathbf{r} \\ + \sum_{\mathrm{B}\in\mathrm{AB}} \sum_j \int \frac{\partial E}{\partial \tilde{\rho}} \frac{\partial \tilde{\rho}}{\partial a_j^{\mathrm{B}}} \frac{\partial a_j^{\mathrm{B}}}{\partial \mathbf{R}_{\mathrm{A}}} d\mathbf{r},\end{aligned} \tag{5.24}$$

which can be simplified with the following notation,

$$\frac{dE(\tilde{\rho})}{d\mathbf{R}_{\mathrm{A}}} = \sum_i \tilde{a}_i^{\mathrm{A}} \frac{\partial I_i^{\mathrm{A}}}{\partial \mathbf{R}_{\mathrm{A}}} + \sum_{\mathrm{B}\in\mathrm{AB}} \mathbf{F}_{\mathrm{AB}} \tag{5.25}$$

where

$$\mathbf{F}_{\mathrm{AB}} = \sum_i \frac{\partial a_i^{\mathrm{A},(\mathrm{AB})}}{\partial \mathbf{R}_{\mathrm{A}}} I_i^{\mathrm{A}} + \sum_j \frac{\partial a_j^{\mathrm{B},(\mathrm{AB})}}{\partial \mathbf{R}_{\mathrm{A}}} I_j^{\mathrm{B}}. \tag{5.26}$$

The first term in Equation 5.25 is evaluated on the grid and is analogues to the GPW expression. The second term is summed over the number of neighbors of A and has no GPW-equivalent. Contributions for this term are only obtained for pairs AB with $\mathrm{A} \notin \mathrm{B}$. "Self-pairs" AA and AA′, where A′ is the periodic image of A, give zero contributions. The derivatives of the expansion coefficients are given by

$$\begin{aligned}\frac{\partial a_i^{\mathrm{A/B}}}{\partial \mathbf{R}_{\mathrm{A}}} = &- \sum_j S_{ij}^{-1} \frac{\partial S_{ji}}{\partial \mathbf{R}_{\mathrm{A}}} S_{ij}^{-1} t_j + \sum_j S_{ij}^{-1} \frac{\partial t_j}{\partial \mathbf{R}_{\mathrm{A}}} \\ &+ \frac{\partial \lambda}{\partial \mathbf{R}_{\mathrm{A}}} \sum_j S_{ij}^{-1} n_j - \lambda \sum_j S_{ij}^{-1} \frac{\partial S_{ji}}{\partial \mathbf{R}_{\mathrm{A}}} S_{ij}^{-1} n_j\end{aligned} \tag{5.27}$$

with

$$\frac{\partial S_{ji}}{\partial \mathbf{R}_{\mathrm{A}}} = \int \frac{\partial f_j^{\mathrm{A}}}{\partial \mathbf{R}_{\mathrm{A}}} f_i^{\mathrm{B}} d\mathbf{r} \tag{5.28}$$

and

$$\frac{\partial t_j}{\partial \mathbf{R}_{\mathrm{A}}} = \sum_{\mu\in\mathrm{A},\nu\in\mathrm{B}} P_{\mu\nu} \frac{\partial}{\partial \mathbf{R}_{\mathrm{A}}} \int \chi_\mu^{\mathrm{A}} \chi_\nu^{\mathrm{B}} f_j^{\mathrm{A/B}} d\mathbf{r}. \tag{5.29}$$

The explicit form of the derivatives of the integrals is discussed below. The derivative of the Lagrange multiplier is

$$\begin{aligned}\frac{\partial\lambda}{\partial\mathbf{R}_\mathrm{A}} = \Bigg[&\left(\frac{\partial N_\mathrm{AB}}{\partial\mathbf{R}_\mathrm{A}} + \mathbf{n}^\mathrm{T}\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial\mathbf{R}_\mathrm{A}}\mathbf{S}^{-1}\mathbf{t} - \mathbf{n}^\mathrm{T}\mathbf{S}^{-1}\frac{\partial\mathbf{t}}{\partial\mathbf{R}_\mathrm{A}}\right)\left(\mathbf{n}^\mathrm{T}\mathbf{S}^{-1}\mathbf{n}\right)\\ &+ \left(N_\mathrm{AB} - \mathbf{n}^\mathrm{T}\mathbf{S}^{-1}\mathbf{t}\right)\left(\mathbf{n}^\mathrm{T}\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial\mathbf{R}_\mathrm{A}}\mathbf{S}^{-1}\mathbf{n}\right)\Bigg]\left[\mathbf{n}^\mathrm{T}\mathbf{S}^{-1}\mathbf{n}\right]^{-2}\end{aligned} \tag{5.30}$$

with

$$\frac{\partial N_\mathrm{AB}}{\partial\mathbf{R}_\mathrm{A}} = \sum_{\mu\in\mathrm{A},\nu\in\mathrm{B}} P_{\mu\nu}\int\frac{\partial\chi_\mu^\mathrm{A}}{\partial\mathbf{R}_\mathrm{A}}\chi_\nu^\mathrm{B}d\mathbf{r}. \tag{5.31}$$

For the derivatives of the expansion coefficients with respect to $\mathbf{R}_\mathrm{B}$ the following symmetry relations hold

$$\frac{\partial a_i^{\mathrm{A},(\mathrm{AB})}}{\partial\mathbf{R}_\mathrm{B}} = -\frac{\partial a_i^{\mathrm{A},(\mathrm{AB})}}{\partial\mathbf{R}_\mathrm{A}} \qquad \text{and} \qquad \frac{\partial a_j^{\mathrm{B},(\mathrm{AB})}}{\partial\mathbf{R}_\mathrm{B}} = -\frac{\partial a_j^{\mathrm{B},(\mathrm{AB})}}{\partial\mathbf{R}_\mathrm{A}}. \tag{5.32}$$

The computation of the stress tensor requires the virial. The contribution to the virial derived from the first term of Equation 5.25 is

$$\Lambda_{kl} = \sum_\mathrm{X}\sum_i \tilde{a}_i^\mathrm{X}\frac{\partial I_i^\mathrm{X}}{\partial h_{kl}}, \tag{5.33}$$

where h_{kl} refers to component k of one of the three lattice vectors; $k,l = x,y,z$. The contribution to the virial related to the second term of Equation 5.25 is given by

$$\Pi_{kl} = \frac{1}{2}\sum_\mathrm{AB}\mathbf{F}_\mathrm{AB}(k)\mathbf{r}_\mathrm{AB}(l), \tag{5.34}$$

where $\mathbf{r}_\mathrm{AB}$ is the vector connecting A and B and $\mathbf{F}_\mathrm{AB}$ is the pair-wise force defined in Equation 5.26. Note that also the periodic "self-pairs" AA′ contribute to the virial since modifying the box size changes the internal interactions between A and its images.

5.2.5. Derivation of auxiliary basis sets

The fitting procedure outlined in the last sections requires an independent set of auxiliary functions $\{f_i\}$ as input. Various approaches have been tested in order to obtain sets of auxiliary functions sufficiently flexible to accurately describe the density, but also sufficiently compact to limit the computational costs.

Generation by geometric progression

The simplest and most successful strategy is to generate uncontracted Gaussian basis sets using a geometric progression for the exponents $\{\tilde{\alpha}_i\}$,

$$\tilde{\alpha}_i = \tilde{\alpha}_\mathrm{min}\cdot q^{i-1} \qquad q = \sqrt[m-1]{\frac{\tilde{\alpha}_\mathrm{max}}{\tilde{\alpha}_\mathrm{min}}}, \tag{5.35}$$

where m is the number of exponents in the auxiliary basis set. The maximal exponent $\tilde{\alpha}_{\max}$ and the minimal exponent $\tilde{\alpha}_{\min}$ are trivially obtained from the exponents of the orbital basis $\alpha_{\max}$ and $\alpha_{\min}$ as

$$\tilde{\alpha}_{\max} = 2\alpha_{\max} \qquad \tilde{\alpha}_{\min} = 2\alpha_{\min}. \tag{5.36}$$

Analogously, the largest angular quantum number is set to $\tilde{l}_{\max} = 2l_{\max}$. Note that higher angular momentums are usually necessary for hydrogen.

The auxiliary basis sets given in Appendix B.2 have all been generated by geometric progression and are tailored for the Molopt family of basis sets [122], which are good quality basis sets for ground state calculations using generalized gradient approximation (GGA) functionals. The Molopt basis sets are compact and highly contracted. For example, the double-ζ basis set for oxygen contains 14 contracted Cartesian Gaussian functions corresponding to 98 primitive functions. In general, the RI basis set has to be 3-4 times larger than the primary basis, which implies that a set of 300-400 functions is required. Indeed, extensive tests showed that smaller LRI basis sets lead to highly inaccurate energies and molecular properties or even to the failure of the fit. In order to reduce the size of the LRI basis sets, the optimization of the LRI exponents has been attempted. This attempt has not been successful and is in the following merely described for the sake of completeness.

Optimization attempts

The LRI fitting approach is not variational and an optimization procedure [52] as proposed for global RI schemes is not possible. Nevertheless, the integrated square of the difference between fitted and exact pair density, the residual D_{AB}, can be minimized employing a global optimizer. Slightly reformulating Equation 5.7 yields

$$D_{\mathrm{AB}} = \int \rho_{\mathrm{AB}}^2 d\mathbf{r} - 2\int \rho_{\mathrm{AB}}\tilde{\rho}_{\mathrm{AB}} d\mathbf{r} + \int \tilde{\rho}_{\mathrm{AB}}^2 d\mathbf{r} \tag{5.37}$$

where

$$\int \rho_{\mathrm{AB}}^2 d\mathbf{r} = \sum_{\mu\in\mathrm{A},\nu\in\mathrm{B}} \sum_{k\in\mathrm{A},l\in\mathrm{B}} P_{\mu\nu}P_{kl} \int \chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_k^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r})\chi_l^{\mathrm{B}}(\mathbf{r}) d\mathbf{r} \tag{5.38}$$

and

$$\begin{aligned} \int \rho_{\mathrm{AB}}\tilde{\rho}_{\mathrm{AB}} d\mathbf{r} &= \sum_{\mu\in\mathrm{A},\nu\in\mathrm{B}} P_{\mu\nu} \int \chi_\mu^{\mathrm{A}}(\mathbf{r})\chi_\nu^{\mathrm{B}}(\mathbf{r}) \left[\sum_i a_i^{\mathrm{A}} f_i^{\mathrm{A}} + \sum_j a_j^{\mathrm{B}} f_j^{\mathrm{B}}\right] \\ &= \sum_i a_i^{\mathrm{A}} t_i^{\mathrm{A}} + \sum_j a_j^{\mathrm{B}} t_j^{\mathrm{B}} \end{aligned} \tag{5.39}$$

and

$$\int \tilde{\rho}_{\mathrm{AB}}^2 d\mathbf{r} = \sum_i \sum_k a_i^{\mathrm{A}} a_k^{\mathrm{A}} S_{ik}^{\mathrm{AA}} + 2\sum_i \sum_l a_i^{\mathrm{A}} a_l^{\mathrm{B}} S_{il}^{\mathrm{AB}} + \sum_j \sum_l a_j^{\mathrm{B}} a_l^{\mathrm{B}} S_{jl}^{\mathrm{BB}}. \tag{5.40}$$

The total residual D is defined as sum over all pair contributions,

$$D(\tilde{\alpha}_i, \tilde{c}_i) = \sum_{\mathrm{AB}} D_{\mathrm{AB}}(\tilde{\alpha}_i, \tilde{c}_i), \tag{5.41}$$

and it is optimized with respect to the exponents $\{\tilde{\alpha}_i\}$ and the contraction coefficients $\{\tilde{c}_i\}$ of the auxiliary basis sets. Note that the first term in Equation 5.37 is a constant and is unaffected by the optimization procedure. The optimization procedure for the LRI basis set has been implemented in CP2K including some optional parameters to limit the range of the exponents. The optimization has been performed for small molecules like O_2 or H_2O usually giving values for D in the range of $10^{-7} - 10^{-6}$ a.u. The residual D is calculated for the converged GPW wavefunction, passed to a global optimizer [215] and re-evaluated at each step. The integrals required for D are calculated analytically allowing for a rather fast update of D. A few hundred optimization steps can be performed per minute on one CPU.

Even though rather small final values of D can be obtained, the resulting basis sets are not flexible enough unless as large and complete as those reported in Appendix B.2. Since all attempts to derive smaller basis sets have failed, the basis sets obtained from the geometric progression have been used for all applications presented in this work.

5.3. Implementation details

5.3.1. General procedure

The LRIGPW approach has been implemented in the DFT part of the CP2K program package [121, 191]. The pseudocode for the LRIGPW implementation is shown in Figure 5.2. In the following, the overlap integrals $\int \chi^{\mathrm{A}}_{\mu}\chi^{\mathrm{B}}_{\nu} d\mathbf{r}$ and $\int f^{\mathrm{A}}_{i} f^{\mathrm{B}}_{j} d\mathbf{r}$ are abbreviated as $(a|b)$ and $(\tilde{a}|\tilde{b})$ and $\int \chi^{\mathrm{A}}_{\mu}\chi^{\mathrm{B}}_{\nu} f^{\mathrm{A}}_{i} d\mathbf{r}$ and $\int \chi^{\mathrm{A}}_{\mu}\chi^{\mathrm{B}}_{\nu} f^{\mathrm{B}}_{j} d\mathbf{r}$ as $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, respectively.

The density has to be fitted at each step of the self-consistent field (SCF) procedure by solving Equation 5.9. However, most of the matrices and vectors needed to solve the linear system of equations or to update the KS matrix are calculated just once when the SCF procedure is initialized. The self-overlap and the vector $\mathbf{n}$ can be pre-calculated for each elemental type, i.e., for each type of LRI basis set. Indeed, the overlap matrix $\mathbf{S}$ and its inverse, the integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ and the derivatives $da^{\mathrm{A}}_i/dP_{\mu\nu}$ and $da^{\mathrm{B}}_j/dP_{\mu\nu}$ do not depend on the density matrix. At each SCF iteration, the vector $\mathbf{t}$ and the Lagrange multiplier λ are easily constructed from the pre-computed integrals and the density matrix. Equation 5.9 is solved and the fitted density is subsequently collocated on the real space grid for each atom X. Once the fitted density is available, the standard GPW procedures are applied to compute the Hartree and XC potentials. The sum of the electrostatic and the XC contribution to the potential is integrated for each atom X yielding the vector elements $\{I^{\mathrm{X}}_i\}$. The

Loop over all atomic kinds K
  Calculate vector elements n_i^K
  Calculate self-overlap S_{ij}^{KK} with $i, j \in$ auxiliary basis set
End K loop
Loop over all pairs AB
  Calculate integrals $(a|b)$ and $(\tilde{a}|\tilde{b})$
  Calculate integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$
  Construct and invert matrix $\mathbf{S}$
  Calculate $\mathbf{n}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{n}$ and $\mathbf{S}^{-1}\mathbf{n}$
  Calculate $\frac{da_i^{\mathrm{A}}}{dP_{\mu\nu}}$ and $\frac{da_j^{\mathrm{B}}}{dP_{\mu\nu}}$
End AB loop
Start SCF cycle
  Calculate density matrix $P_{\mu\nu}$
  Loop over all pairs AB
    Calculate $\mathbf{t}$ and Lagrange multiplier λ
    Get expansion coefficients for AB: $\mathbf{a} = \mathbf{S}^{-1}\mathbf{t} + \lambda\mathbf{S}^{-1}\mathbf{n}$
    Sum up $\tilde{a}_i^{\mathrm{A}}$: $\tilde{a}_i^{\mathrm{A}} = \tilde{a}_i^{\mathrm{A}} + a_i^{\mathrm{A,(AB)}}$
    Sum up $\tilde{a}_j^{\mathrm{B}}$: $\tilde{a}_j^{\mathrm{B}} = \tilde{a}_j^{\mathrm{B}} + a_j^{\mathrm{B,(AB)}}$
  End AB loop
  Loop over all atoms X
    Collocate $\tilde{\rho}_{\mathrm{X}}(\mathbf{r}) = \sum_i \tilde{a}_i^{\mathrm{X}} f_i^{\mathrm{X}}(\mathbf{r})$ on the real space grid
    Sum up total density: $\tilde{\rho}(\mathbf{r}) = \tilde{\rho}(\mathbf{r}) + \tilde{\rho}_{\mathrm{X}}(\mathbf{r})$
  End X loop
  Transfer to reciprocal space: $\tilde{\rho}(\mathbf{r}) \xrightarrow{\mathrm{FFT}} \tilde{\rho}(\mathbf{G})$
  Solve Poisson equation: $\tilde{\rho}(\mathbf{G}) \longrightarrow \tilde{V}_{\mathrm{H}}(\mathbf{G})$
  Transfer back to real space: $\tilde{V}_{\mathrm{H}}(\mathbf{G}) \xrightarrow{\mathrm{FFT}^{-1}} \tilde{V}_{\mathrm{H}}(\mathbf{r})$
  Calculate $\tilde{V}_{\mathrm{xc}}$ and sum up: $\tilde{V}(\mathbf{r}) = \tilde{V}_{\mathrm{H}}(\mathbf{r}) + \tilde{V}_{\mathrm{xc}}(\mathbf{r})$
  Loop over all atoms X
    Integrate potential: $I_i^{\mathrm{X}} = \int \tilde{V}(\mathbf{r}) f_i^{\mathrm{X}}(\mathbf{r}) d\mathbf{r}$
  End X loop
  Loop over all pairs AB
    Calculate KS matrix: $\tilde{H}_{\mu\nu} = \sum_i \frac{da_i^{\mathrm{A}}}{dP_{\mu\nu}} I_i^{\mathrm{A}} + \sum_j \frac{da_j^{\mathrm{B}}}{dP_{\mu\nu}} I_j^{\mathrm{B}}$
  End AB loop
  IF SCF converged exit
End SCF cycle

Figure 5.2.: Pseudocode for LRIGPW for a single point calculation. Only operations directly related to LRIGPW are displayed. Evaluation of other terms, e.g. pseudopotentials, are as in GPW and not reported.

KS matrix is finally updated for each pair from the pre-calculated derivatives of the expansion coefficients and the vector elements $\{I_i^{\mathrm{X}}\}$.

If forces are required, the derivatives of the integrals $(a|b)$, $(\tilde{a}|\tilde{b})$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ with respect to the positions are necessary, see Equations 5.25 and 5.26. $\mathbf{F}_{\mathrm{AB}}$ is then built from the pre-computed matrices and their derivatives.

The LRIGPW operations shown in Figure 5.2 can be easily parallelized using a standard message passing interface (MPI). The computation of the LRI integrals, the evaluation of the expansion coefficients and the update of the KS matrix are performed independently for each pair. These tasks are parallelized by distributing the pairs over the processors. For the grid-based collocation of the density and the integration of the potential, the distribution is applied to the grid points.

5.3.2. Evaluation of LRI integrals

The LRI integrals $(a|b)$, $(\tilde{a}|\tilde{b})$ $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ are overlap integrals of contracted spherical Gaussian functions that are calculated exploiting efficient recursion schemes. The most commonly used approach is the Obara-Saika (OS) recursion [216], which is applied to primitive Cartesian Gaussian functions. The transformation from Cartesian to spherical functions and the contraction steps are performed on top of the already computed integrals. Integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ include two functions located at the same center. The product of two Cartesian Gaussians centered at the same atom gives a new Gaussian,

$$\begin{aligned}\phi(\alpha,\mathbf{l},\mathbf{r},\mathbf{R})\tilde{\phi}(\tilde{\alpha},\tilde{\mathbf{l}},\mathbf{r},\mathbf{R}) = \\ (x-R_x)^{l_x+\tilde{l}_x}(y-R_y)^{l_y+\tilde{l}_y}(z-R_z)^{l_z+\tilde{l}_z}\exp\left[-(\alpha+\tilde{\alpha})(\mathbf{r}-\mathbf{R})^2\right]\end{aligned} \tag{5.42}$$

with exponent $(\alpha+\tilde{\alpha})$ and angular momentum $L = l_x+\tilde{l}_x+l_y+\tilde{l}_y+l_z+\tilde{l}_z$. The $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ integrals are then calculated like ordinary two-center overlap integrals. The angular momentum $L_{\max}$ is the sum of $l_{\max}$ (primary basis) and $\tilde{l}_{\max}$ (auxiliary basis set). Typical values of $L_{\max}$ are between 6 to 8 for the basis sets used. The OS recursion for such large angular momentums is rather expensive. Instead, the integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ can be formally computed as three-center overlap integrals, which is generally less efficient. However, the recursion is in that case performed for smaller angular momentums and turned out to be slightly more efficient.

Calculating $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ becomes computationally expensive for periodic systems. Using the Molopt basis set in combination with the auxiliary basis sets given in Appendix B.2, the dimensions of the integral matrix computed for Cartesian primitives is in the range of $100 \times 100 \times 400$ taking the oxygen example of Section 5.2.5. The recursion and contraction operations have to be carried out for each pair which makes this part of the calculation the dominating step of the electronic structure optimization. The situation becomes even worse when the derivatives of these integrals

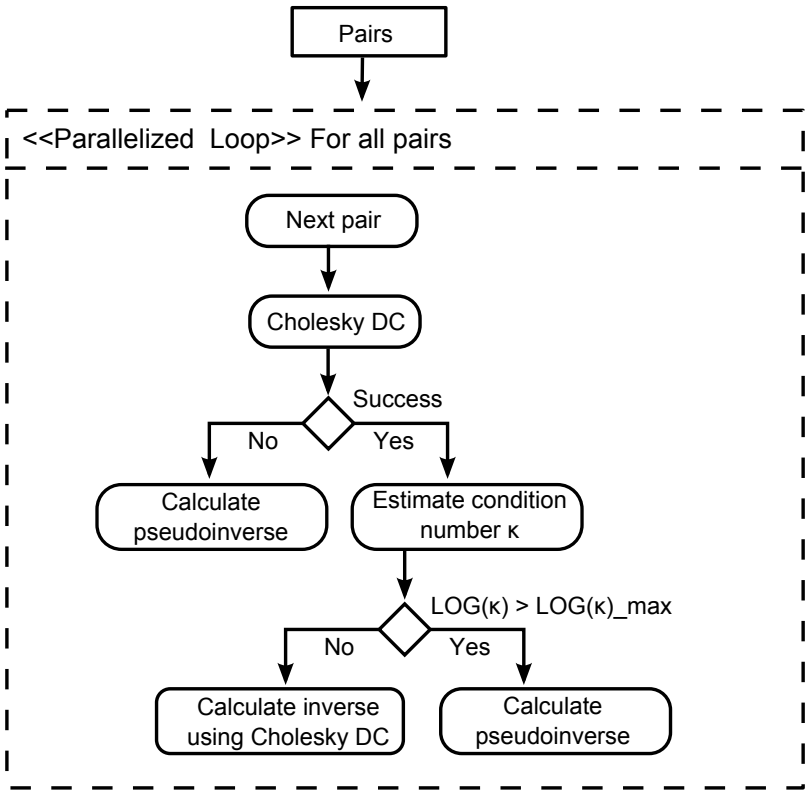

Figure 5.3.: Flow chart for calculating the inverse or pseudoinverse of **S**. Cholesky DC is short for Cholesky decomposition.

are required since the derivatives are constructed from higher angular functions [216] and the contraction has to be performed for each spatial derivative separately.

An alternative analytical integration scheme employing solid harmonic Gaussian (SHG) functions has been implemented in order to reduce the computational cost for these integrals. The main advantage of the proposed method is the reduced number of operations required for the contraction step. Furthermore, recursive elements are pre-computed and re-used multiple times. The SHG integral scheme is described in Chapter 6.

5.3.3. Inversion of the overlap matrix

For very large auxiliary basis sets, the overlap matrix **S** becomes ill-conditioned, i.e., the matrix is numerically positive semi-definite. The inversion of an ill-conditioned matrix **S** is numerically unstable and leads to SCF convergence problems. This problem can be solved by calculating the pseudoinverse instead. For the pseudoinverse, a singular value decomposition (SVD) of **S** is performed. Singular values smaller than a given threshold are skipped and the pseudoinverse is constructed following the method

proposed in Ref. [217]. Since **S** is quadratic and symmetric, the SVD can be replaced by a diagonalization, which is, however, not much more efficient. The calculation of the pseudoinverse is computationally 10-20 times more expensive than a simple inversion and might significantly affect the performance. Therefore, the procedure displayed in Figure 5.3 has been implemented. The SCF usually converges without problems, even when the condition number κ of matrix **S** is rather large. Convergence problems are in general attributed to a few pairs that have overlap matrices with extremely large condition numbers ($\log \kappa > 20$). Only for these problematic pairs the pseudoinverse algorithm is adopted, thus limiting the computational effort. The condition number is estimated from the Cholesky decomposition of **S**. When $\log \kappa$ is larger than a predefined threshold, the pseudoinverse is preferred to standard inversion. This happens only for a small fraction of the pairs ($\sim$1%). When using the pseudoinverse, the expression of the forces should also be adapted. However, the numerical inconsistencies introduced by keeping the same expression already used for the standard inversion appears to be negligible according to our tests.

The condition number of **S** is indeed rather large using the auxiliary basis sets given in Appendix B.2. Nevertheless, computing the inverse of **S** (for all pairs) is sufficient for most systems. The SCF is typically stable enough to converge to a target accuracy of 10^{-6} a.u. with respect to changes in the density matrix $P_{\mu\nu}$. Computing the pseudoinverse following the procedure in Figure 5.3 is only carried out when necessary, e.g., for problematic systems or when tighter convergence criteria are set. The regular inversion of **S** is preferred due to the lower computational cost, but also because it is more accurate. Namely, by computing the pseudoinverse we introduce another approximation on top of the density fitting. By increasing the threshold for the accepted singular values, the numerical stability is improved. On the other hand, the fitting becomes less flexible and it becomes increasingly difficult to reproduce the GPW density. In our setup, singular values smaller than 10^{-8} are usually removed, which greatly improves the stability, but reduces the accuracy of the fit only slightly.

5.4. Computational details

5.4.1. General setup

LRIGPW and GPW calculations are carried out for isolated molecules and dimers as well as periodic structures, such as ice XV, molecular crystals, graphene and liquid water. All calculations are performed with the CP2K program package at the KS-DFT level. For the primary basis sets, double-ζ plus polarization (DZVP) and triple$-\zeta$ plus double polarization (TZV2P) basis sets of the Molopt type [122] are used. The LRIGPW tests are performed with medium- (m-aux) and large-sized (l-aux) auxiliary basis sets, see Appendix B.2. Both auxiliary basis sets contain the same number of exponents, but the large basis sets include additionally functions with higher angular

Table 5.1.: PW cutoffs for all systems.

system	cutoff [Ry]	relative cutoff [Ry]
isolated molecules, $(HF)_2$ dimer	400	40
$(NH_3)_2$, $(H_2O)_2$, formic acid dimer	600	40
graphene	500	40
GMTKN24 benchmark	600	40
ice XV	800	50
molecular crystals	800	60
H_2O boxes (benchmark)	400	50
$128H_2O$ box (MD)	300	50

momentum for the large exponents. For hydrogen, an alternative auxiliary basis (m^+-aux) with more diffuse functions is also available. If not indicated otherwise, the valence electrons are described at the DZVP level and the m-aux basis set is used. Norm-conserving pseudopotentials [123–125] are employed to approximate the core-electron interactions. The Perdew-Becke-Ernzerhof (PBE) [20] functional is employed for the XC potential. Unless otherwise stated, the Grimme D3 correction [34] is employed to model dispersion interactions. The PW cutoff has been adjusted for each type of system individually and is reported in Table 5.1. The optimization of the molecular orbitals is achieved with the orbital transformation (OT) technique [218] for all systems, except graphene. In the latter case, a simple diagonalization of the KS matrix is used. Periodic boundary conditions are applied in all spatial directions for condensed matter systems. For graphene, a slab model with 25Å vacuum is used to avoid spurious interactions with the periodic images.

5.4.2. Systems and properties investigated

Dimers The accuracy of LRIGPW for intermolecular distances is studied for the hydrogen-bonded dimers of HF, NH_3, H_2O and formic acid. The ammonia dimer has a very flat potential energy surface and several minima exist. The ammonia dimer considered here has D_{2h} symmetry and is referred to as "cyclic" in Ref. [219]. The structure optimizations of the dimers are not counterpoise (CP) corrected.
Graphene The equilibrium lattice constant is calculated using a supercell of 12×12 units.
GMTKN24 benchmark suite Reaction energies are studied for the WATER27, DARC and S22 subsets of the GMTKN24 benchmark database [220]. The WATER27 subset assesses the decomposition energies of neutral $(H_2O)_n$, positively $H^+(H_2O)_n$ and negatively charged $OH^-(H_2O)_n$ water clusters. The DARC subset is a collection of different Diels-Alder reactions. S22 includes binding energies of hydrogen-bonded

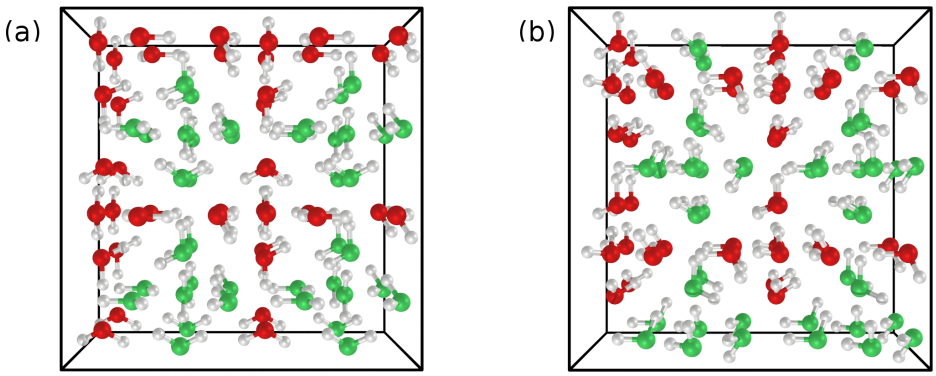

Figure 5.4.: Two of the possible 18 symmetry inequivalent structures of ice XV. In both cases, 2×2×2 supercells are shown along the [001] axis. (a) 2C1 structure with $P\bar{1}$ symmetry (b) 9A2 structure exhibiting Cc symmetry. Color code: H, white; O, green and red indicating distinct sublattices.

dimers as well as complexes with predominant London dispersion interactions. For the labeling of the reactions see Appendix B.3. These PBE calculations are carried out without D3 correction. For DARC and S22, the m$^+$-aux basis set is used for hydrogen to provide more flexibility needed to model complexes with difficult electronic structure. The relative difference between the LRIGPW and GPW reaction energies is calculated as

$$\Delta_{\text{rel}}^{\text{GPW}} = \frac{|\Delta E_{\text{GPW}} - \Delta E_{\text{LRI}}|}{\Delta E_{\text{GPW}}}. \tag{5.43}$$

The relative errors with respect to the CCSD(T) reference value [220] are defined as,

$$\Delta_{\text{rel}}^{\text{CCSD(T)}} = \frac{|\Delta E_{\text{PBE}} - \Delta E_{\text{CCSD(T)}}|}{\Delta E_{\text{CCSD(T)}}}. \tag{5.44}$$

Ice XV is the latest discovered polymorph of ice [221–224]. It is one of the 15 known distinct crystalline ice phases and can be regarded as the hydrogen-ordered counterpart of ice VI. Ice XV is one of the high-pressure ice phases and occurs in the ∼1 GPa regime. Starting from ice VI, the hydrogen network can order in three different ways labeled as A, B and C. Each of them exhibits two distinct image forms. These six subunits can be combined to 18 symmetry inequivalent structures of ice XV with $P\bar{1}$ (#2), $P2_1$ (#4), Pn (#7) and Cc (#9) space group symmetry. Within this notation, 4C2 refers, for example, to one of the two possible $P2_1$ structures with C-type ordering. For more details on the labeling conventions see Ref. [223, 224]. In Figure 5.4, the experimentally determined 2C1 structure [221] and the theoretically predicted 9A2 structure [224] are shown. Both structures consist of two interpenetrating, polar sublattices. Due to the $P\bar{1}$ symmetry, the dipole moments of both sublattices cancel for the antiferro-electric 2C1 structure. On the contrary, 9A2 has a net dipole and is ferro-electric.

The unit cell of ice XV consists of 10 water molecules and has been extended to a 2×2×2 supercell. The structures and the supercell parameters have been retrieved

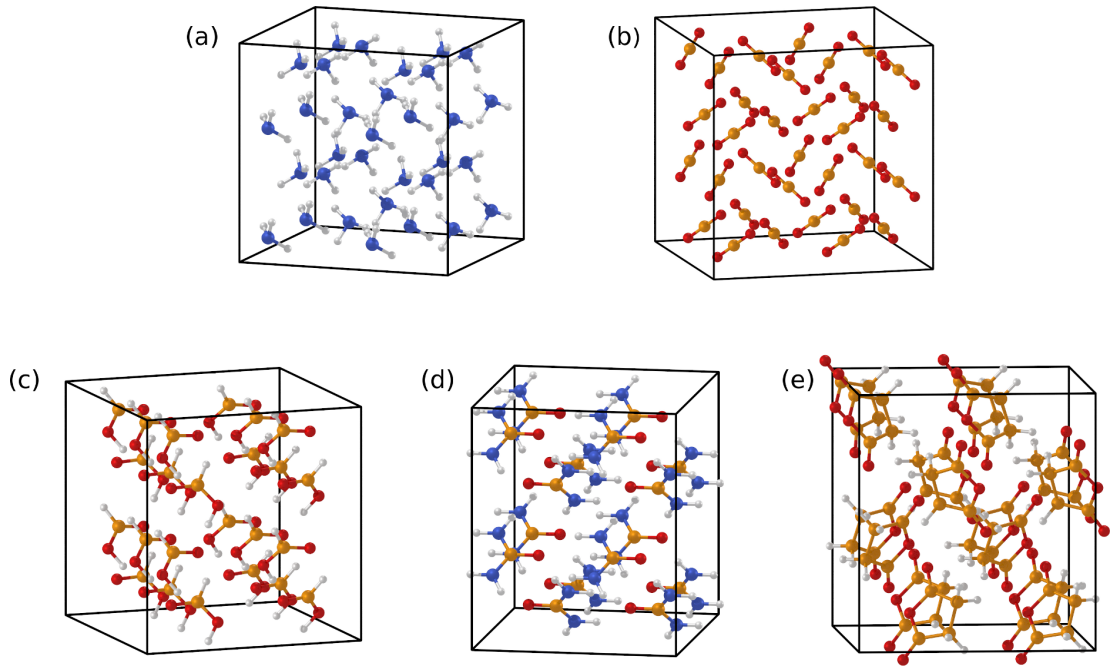

Figure 5.5.: Supercells of molecular crystals studied and number of molecules N_{mol} per cell. (a) NH_3, 2×2×2 supercell, N_{mol}=32. (b) CO_2, 2×2×2 supercell, N_{mol}=32 (c) Formic Acid, 1×3×2 supercell, N_{mol}=24 (d) Urea, 2×2×2 supercell, N_{mol}=16 (e) Succinic anhydride, 2×2×1 supercell, N_{mol}=16. Color code: C orange, O red, N blue, H white.

from the Supporting Information (SI) of Ref. [224]. In the original work, the geometry and the volume were relaxed under 1 GPa isotropic external pressure at the RI-Møller-Plesset (RI-MP2) level of theory using a cc-TZ basis set. The relaxed cells are triclinic slightly deviating from an orthorhombic configuration. Starting from the relaxed structures and cell vectors, molecular and total dipole moments are computed from the center of the atoms and the centers of the maximally localized Wannier orbitals (MLWO) [131, 132]. The Berry-Phase approximation [225] is also used to estimate net dipole moments. Infrared (IR) spectra are obtained from normal mode analysis by calculating the variations of the dipole moments along the vibrations.

Molecular crystals Cohesive energies and lattice parameter of five different molecular crystals are studied: NH_3, CO_2, formic acid, urea and succinic anhydride. Their geometries and supercells are displayed in Figure 5.5. The experimental structures of the formic acid, urea and succinic anhydride crystals have been retrieved from the Cambridge Structural Database [226]. The respective reference codes are FORMAC01, UREAXX09 and SUCANH15. The structures of NH_3 and CO_2 are obtained from work [54] building the geometries from the experimental lattice parameters and the information on the space group. The geometries and lattice vectors have been relaxed at the GPW level starting from the experimental cell parameters given in Table 5.5. The cohesive energies are computed for the relaxed structures as suggested

in Refs. [54, 227]. The cohesive energy at a given volume V is defined as

$$E_{\mathrm{coh}}(V) = \frac{E_{\mathrm{supercell}}(V)}{N_{\mathrm{mol}}} - E_{\mathrm{mol}}^{\mathrm{gas}} - E_{\mathrm{CP}}, \tag{5.45}$$

where N_{mol} is the number of molecules in the supercell, $E_{\mathrm{supercell}}(V)$ is the total energy of the supercell and $E_{\mathrm{mol}}^{\mathrm{gas}}$ is the gas phase energy of an isolated, relaxed molecule. The CP correction E_{CP} is defined as

$$E_{\mathrm{CP}} = E_{\mathrm{mol+ghost}}^{\mathrm{crystal}}(V) - E_{\mathrm{mol}}^{\mathrm{crystal}}(V), \tag{5.46}$$

where $E_{\mathrm{mol+ghost}}^{\mathrm{crystal}}(V)$ and $E_{\mathrm{mol}}^{\mathrm{crystal}}(V)$ are the energies of one single molecule in the geometry it takes in the crystal, with and without ghost functions. For NH_3 and CO_2, the first two coordination shells, i.e. the 12 nearest neighbors, are considered as ghost atoms, whereas for the other molecular crystals only atoms of the first coordination shell are included.

Liquid Water Benchmark calculations are performed for liquid water boxes increasing the size from 32 to 512 water molecules. The liquid water boxes are orthorhombic and the box lengths are adjusted such that the density is always 1.0 g/cm^3.

MD simulations of liquid water are carried out within the canonical (NVT) ensemble for the water box with 128 molecules. The temperature is set to 330 K and the time step is 0.5 fs. The system has been equilibrated for about 4 ps. Structural features are sampled from, at least, the last 25 ps of the simulation.

5.5. Results and discussion

5.5.1. Total energies and structure parameters

Differences between exact and fitted density are usually very small unless the fit breaks down, e.g., due to a too small auxiliary basis set. The total residual $\sum_{\mathrm{AB}} D_{\mathrm{AB}}$ is usually in the range of $10^{-7} - 10^{-6}$ a.u. and is not a good measure for the accuracy. The error caused by the LRI approach is better assessed by comparing molecular energies and structure parameters.

Total energies

Total energies of a few small isolated molecules and of liquid water boxes are compared to the GPW value. The difference between the absolute values are reported per atom in Table 5.2. The single point energies are obtained with different combinations of the DZVP and TZV2P primary basis sets and the medium- and large-sized fit sets. For the single molecules, total energies are given for structures relaxed at the GPW-DZVP level. The LRI error in the total energy is mostly less than 100 μE_{H} ($\sim$0.26kJ/mol), for DZVP combined with the m-aux basis set. Employing a larger auxiliary basis sets,

Table 5.2.: Differences in total energies ΔE (μE_H) per atom comparing GPW and LRIGPW using two different primary and two different auxiliary basis sets. ΔE calculated for isolated molecules and liquid water (liq) boxes of different sizes.

	DZVP		TZV2P	
	ΔE(m-aux)	ΔE(l-aux)	ΔE(m-aux)	ΔE(l-aux)
HF	263	66	258	70
HCl	72	75	71	77
O_2	1	74	6	72
CO_2	90	2	191	99
H_2O	112	59	116	56
NH_3	71	60	75	64
CH_4	89	76	86	71
Borazine*	497	–	–	–
$(32H_2O)_{liq}$	41	62	190	202
$(128H_2O)_{liq}$	68	35	51	157
$(512H_2O)_{liq}$	27	57	164	188

*l-aux and TZV2P basis sets not available for B.

all errors are smaller than 80 μE_H. The TZV2P basis set captures more features of the electronic structure and should be more difficult to fit. Nevertheless, the deviations are only slightly larger. The LRI errors are in the range reported for global RI schemes applied to the Coulomb part, where total energies are affected by errors of less than 200 μE_H [45, 70]. Further reduction of the error to less than 60 μE_H has been reported with carefully optimized auxiliary basis sets [71]. However, errors larger than 1000 μE_H have also been reported for critical cases [70, 71]. Such large errors have not been observed with LRIGPW so far.

The errors in the total energies are generally smaller when using a larger auxiliary basis set. For some cases, the opposite is found. This must be attributed to the fact that the numerical stability is lower for the larger auxiliary basis sets causing numerical noise of several μE_H. The reason for the latter is that the condition number of the overlap matrix $\mathbf{S}$ is increased, see Section 5.3.3. Total energies can only be a first hint to evaluate the accuracy of LRIGPW. More relevant are relative energies and structure parameters. For such properties, the accuracy is indeed always improved when increasing the size of the auxiliary basis set, as shown below.

Intramolecular structure parameters

In Table 5.3, intramolecular distances and angles are presented for small molecules. The optimizations are carried out at the DZVP and TZV2P level employing the m-

Table 5.3.: Intramolecular geometry parameters obtained with two different orbital basis sets. For LRIGPW, the m-aux basis sets have been employed. Bond lengths d are given in Å and angles θ in degrees.

		DZVP			TZV2P		
	quantity	GPW	LRIGPW	Δ	GPW	LRIGPW	Δ
HF	d_{FH}	0.9379	0.9377	0.0002	0.9315	0.9314	0.0001
HCl	d_{ClH}	1.2898	1.2899	0.0001	1.2877	1.2877	0.0000
HBr*	d_{BrH}	1.4384	1.4383	0.0001	–	–	–
O_2	d_{OO}	1.2308	1.2304	0.0004	1.2270	1.2264	0.0006
CO_2	d_{OC}	1.1776	1.1777	0.0001	1.1756	1.1754	0.0002
	θ_{OCO}	180.00	180.00	0.00	180.00	180.00	0.00
H_2O	d_{OH}	0.9719	0.9717	0.0002	0.9701	0.9701	0.0000
	θ_{HOH}	104.04	104.18	0.14	104.29	104.39	0.10
NH_3	d_{NH}	1.0253	1.0254	0.0001	1.0214	1.0217	0.0003
	θ_{HNH}	105.84	105.91	0.07	106.35	106.44	0.09
CH_4	d_{CH}	1.0962	1.0962	0.0000	1.0954	1.0955	0.0001
	θ_{HCH}	109.47	109.47	0.00	109.47	109.47	0.00

*TZV2P basis set not available for Br

aux basis set for LRI. For both primary basis sets, the LRI errors are in the same range. The error in the bond lengths is smaller than 0.6 mÅ and the angles deviate by less than 0.14°. For comparison, the global RI errors are reported to be smaller than 1 mÅ for the bond length and 0.1° for the angles [45, 70, 71].

We can argue that we have almost reached the basis set limit with the TZV2P basis sets. Comparing the GPW difference between DZVP and TZV2P is thus a measure for the incompleteness of the DZVP basis set. In fact, the bond length and angle differences between GPW structures optimized with the two different primary basis sets are up to 6 mÅ and 0.5°, respectively. The error introduced by the small primary basis set is thus significantly larger than the error of the fit.

Intermolecular structure parameters

For the simulation of liquid phase systems or condensed phase systems in general, the accurate description of intermolecular interactions is very important. In order to evaluate the LRIGPW accuracy for intermolecular properties, the geometries of the HF, NH_3, H_2O and formic acid dimers are compared to GPW. The dimers are optimized at the DZVP and TZV2P level. The dimer geometries are displayed in Figure 5.6 and the intermolecular structure parameters, such as internuclear distances, angles and dihedral angles, are given in Table 5.4. The interactions between these

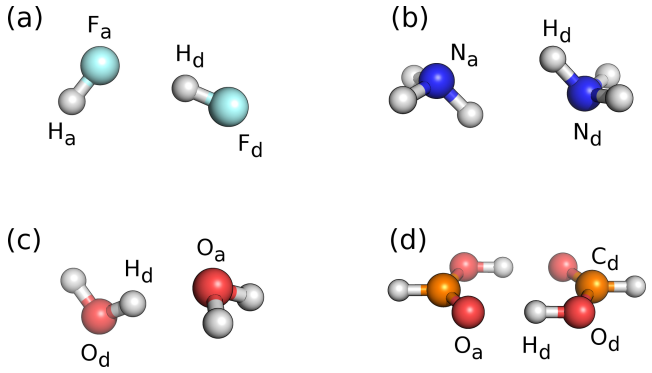

Figure 5.6.: Dimer geometries. (a) HF dimer, (b) NH_3 dimer, (c) H_2O dimer and (d) formic acid dimer. Atoms of the H-bond acceptor molecule are labeled with *a*, atoms of the donor molecule with *d*. Color code: F cyan, H white, N blue, O red, C orange.

Table 5.4.: Intermolecular geometry parameters for the $(HF)_2$, $(NH_3)_2$, $(H_2O)_2$ and formic acid $(FA)_2$ dimers obtained with two different primary basis sets. For LRIGPW, the m-aux basis set is used (except: m^+-aux for $H/(H_2O)_2$ at TZV2P level). Distances R given in Å, angles θ and dihedral angles Ψ in degrees. Site names and labels are specified in Figure 5.6.

		DZVP			TZV2P		
	quantity	GPW	LRIGPW	Δ	GPW	LRIGPW	Δ
$(HF)_2$	$R(F\cdots F)$	2.762	2.765	0.003	2.762	2.766	0.004
	$R(F_a\cdots H_d)$	1.822	1.826	0.004	1.828	1.833	0.005
	$\theta_{H_dF_dF_a}$	5.56	5.66	0.10	5.55	5.80	0.25
	$\theta_{H_aF_aH_d}$	112.71	112.81	0.10	112.29	112.58	0.29
$(NH_3)_2$	$R(N\cdots N)$	3.157	3.154	0.003	3.160	3.158	0.002
	$\theta_{H_dN_dN_a}$	42.29	42.36	0.07	42.22	42.25	0.03
$(H_2O)_2$	$R(O\cdots O)$	2.923	2.928	0.005	2.922	2.917	0.005
	$\theta_{H_dO_dO_a}$	5.26	5.10	0.16	5.03	3.79	1.24
	θ^*_{dip}	120.37	120.30	0.07	119.50	118.26	1.24
$(FA)_2$	$R(C\cdots C)$	3.780	3.735	0.045	3.783	3.728	0.055
	$\theta_{O_aH_dO_d}$	178.75	179.60	0.85	178.53	179.63	1.1
	$\Psi_{O_aH_dO_dC_d}$	180.83	182.08	1.25	179.30	180.13	0.83

$^*\theta_{dip}$: angle between OO-axis and bisection line of the HOH angle of the H-bond acceptor water molecule.

dimers are mediated by hydrogen bonds. Linear hydrogen bonds are found for HF, H_2O and the formic acid dimer, whereas the ammonia dimer is bound non-linearly. The geometry parameters obtained for $(HF)_2$, $(H_2O)_2$ and $(NH_3)_2$ in this work are in good agreement with previously published results [219, 228]. The formic acid dimer is different from the other dimers in the sense that it is planar and exhibits two equivalent linear hydrogen bonds [229]. Therefore, it will be discussed separately.

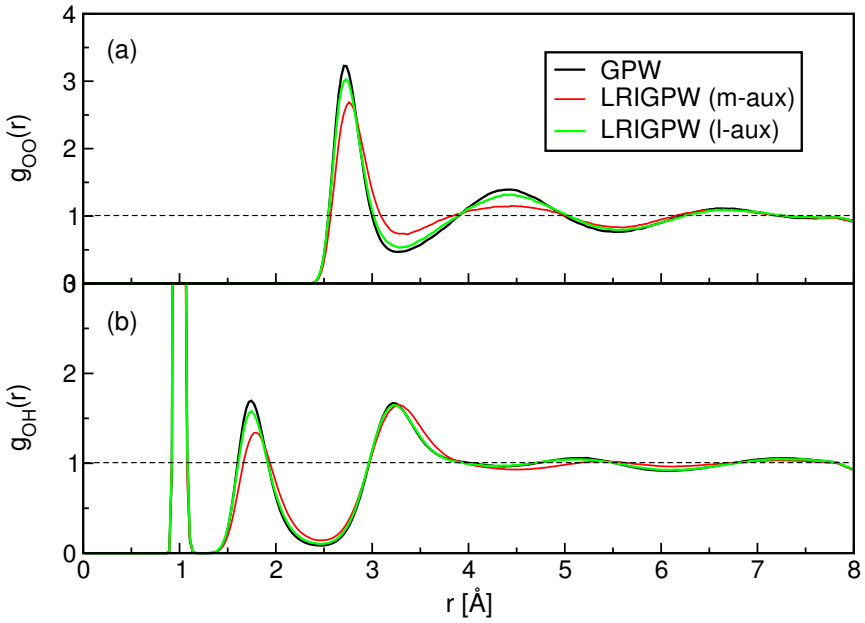

Figure 5.7.: RDFs for liquid water obtained with GPW and LRIGPW at the DZVP level. (a) Oxygen-oxygen and (b) oxygen-hydrogen RDFs.

For $(HF)_2$, $(H_2O)_2$ and $(NH_3)_2$, the LRI internuclear distances computed at the DZVP level are affected by errors smaller than 5 mÅ and the error of the LRI angles is smaller than 0.16°. The differences between the GPW structures computed with the DZVP and TZV2P basis sets are up to 6 mÅ and 0.87° for distances and angles, respectively. Also here we can argue that the error margins due to the incompleteness of the DZVP basis set are larger than the error of the fit. At the TZV2P level, the LRI errors are in most cases not significantly larger. However, the features of the electronic density are more sophisticated and larger auxiliary basis sets are required for a few cases. For example, the m-aux basis set appeared to be insufficient for $(H_2O)_2$ yielding errors in the angles of several degrees. The correct geometry is reproduced by employing the m^+-aux basis, which contains more diffuse functions, for hydrogen.

The planar geometry of the formic acid dimer is preserved using the LRI approach, but the C$\cdots$C distance is underestimated by 45 mÅ which is an order of magnitude larger than for the other dimers. To assess the significance of this deviation, we compare to the error introduced by neglecting the CP correction for the geometries. In work [219], intermolecular distances in dimers are reported to change upon CP correction by as much as 25 mÅ. Assuming that the basis set superposition error (BSSE) is in the same range for the formic acid dimer, we can conclude that an LRI error of 45 mÅ is still acceptable when neglecting the CP correction anyway. Furthermore by including more diffuse fit functions, the LRI error in the C$\cdots$C distance can be also reduced to values smaller than 10 mÅ.

Conclusions that can be drawn from dimer geometries are somewhat limited since no averaging over a larger number of structures is possible. Structural properties ob-

tained from MD trajectories are a much more comprehensive test. Figure 5.7 displays the radial distribution functions (RDFs) sampled over the NVT trajectories of the water box with 128 H_2O molecules. The RDFs obtained with LRIGPW in combination with the m-aux basis sets differ noticeably from the GPW result. The lower and smoother RDF profile indicates that the liquid is less structured, i.e., the intermolecular interactions (hydrogen bonds) are weaker. On the contrary, the GPW-RDFs are almost perfectly reproduced with the larger auxiliary basis set. The remaining differences can be attributed to the different length of the trajectories. The GPW trajectory has been extended for 50 ps, whereas the "LRIGPW (l-aux)" trajectory is only 25 ps long.

5.5.2. Reaction energies

Decomposition and reaction energies are calculated for the WATER27, DARC and S22 subsets of the GMTKN24 benchmark suite. The relative and absolute errors with respect to the GPW results for the WATER27 subset are plotted in Figure 5.8 for two different combinations of primary and auxiliary basis sets. The relative LRI error $\Delta_{\mathrm{rel}}^{\mathrm{GPW}}$ is typically less than 2% and never larger than 5% using DZVP in combination with the m-aux basis sets, see Figure 5.8(a). The agreement is quite good already at this level of approximation, in particular considering that a 5% deviation corresponds to an absolute error of only 3.6 kJ/mol, see Figure 5.8(c). Employing a larger auxiliary basis set, the decomposition energies are almost perfectly reproduced. The maximal deviation with respect to GPW is reduced to 1.3%, see Figure 5.8(b).

One of the largest deviations at the DZVP level (13.1 kJ/mol) is observed for the decomposition energy of the $(H_2O)_{20}$ cluster, reaction 14, see Figure 5.8(c). The absolute difference between the GPW results obtained with the DZVP and TZV2P basis sets is 27.8 kJ/mol for this reaction. Assuming that the basis limit is reached with the TZV2P basis set, we conclude that the LRI error is significantly smaller than the basis set error.

The relative errors $\Delta_{\mathrm{rel}}^{\mathrm{CCSD(T)}}$ with respect to the CCSD(T) reference values are also computed, using both the DZVP and the TZV2P primary basis sets. DFT performs reasonably well for neutral and positively charged water clusters, but it fails for the negatively charged clusters. As expected, GPW and LRIGPW show an equivalent behavior in comparison to CCSD(T), see Figure 5.8(d,e).

For the DARC subset, we find that pure DFT is insufficient as soon as aromatic compounds are involved in the reaction, see Figure 5.8(f). Regarding the fitting, LRIGPW reproduces again the GPW energies very well. The S22 subset includes the binding energy of several hydrogen bonded dimers such as NH_3 and H_2O, for which DFT yields accurate values and LRIGPW works smoothly, see Figure 5.8(g). For the complexes, where dispersion interactions are predominant, DFT fails drastically. However, the fitting remains stable except for reaction 15, which labels a stacked

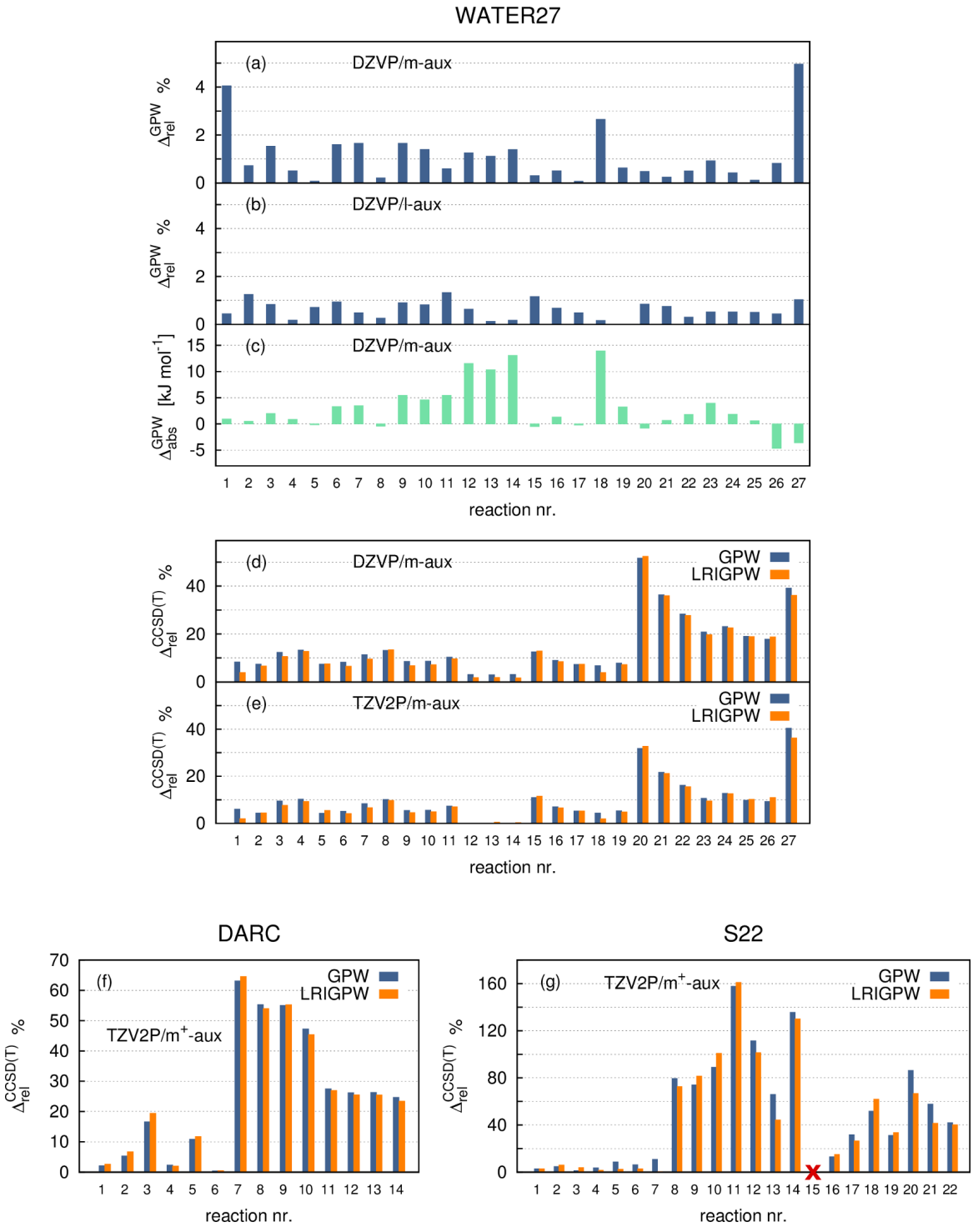

Figure 5.8.: Decomposition/reaction energies for the GMTKN24 subsets WATER27, DARC and S22. Relative errors with respect to GPW $\Delta_{\text{rel}}^{\text{GPW}}$ at the DZVP level for WATER27 using the (a) m-aux and (b) l-aux basis sets. (c) Absolute errors $\Delta_{\text{abs}}^{\text{GPW}}$ with respect to GPW. Relative errors $\Delta_{\text{rel}}^{\text{CCSD(T)}}$ with respect to CCSD(T) for WATER27 at (d) the DZVP and (e) TZV2P level and for (f) DARC and (g) S22 at the TZV2P level.

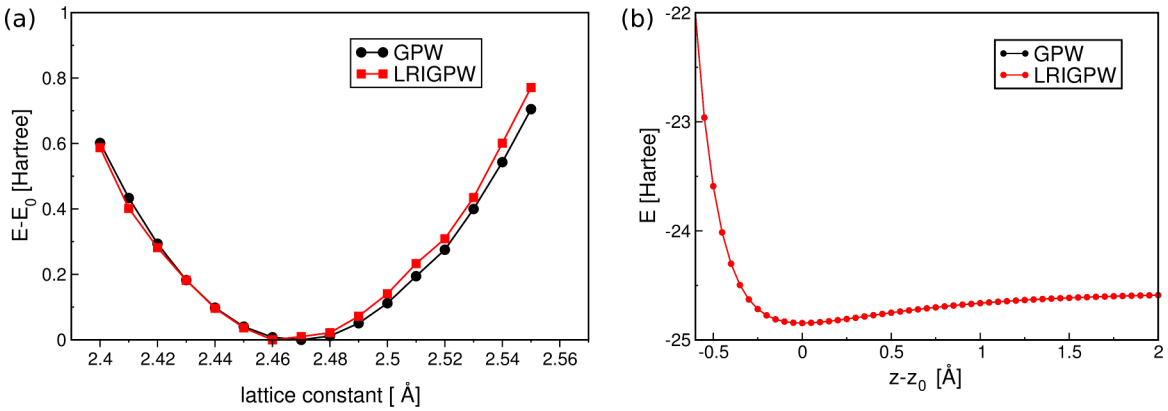

Figure 5.9.: (a) Lattice optimization for a graphene sheet. Displayed are the relative energies. E_0 is the total energy at the equilibrium lattice constant. (b) Dissociation curve of HF. z indicates the H-F distance and z_0 the equilibrium distance.

conformation of two DNA bases. In this case, a break down of the LRI approach has been observed using TZV2P as primary basis. At the DZVP level instead, LRIGPW is stable for this reaction, deviating from GPW by 1.8 kJ/mol. These results suggest that the m-aux basis sets are not always flexible enough to capture the detailed features of the electronic structure at the TZV2P level.

5.5.3. Graphene lattice and HF dissociation curve

In order to test the performance of LRIGPW for 2D periodic materials, a lattice parameter optimization of graphene is carried out. The total energy of the graphene layer has been evaluated by varying progressively the lattice parameter from 2.40 to 2.56 Å, and accordingly re-scaling the coordinates. The energy has been plotted dependent on the lattice constants, showing that the equilibrium is reached for a_0 between 2.46 and 2.47 with both methods, see Figure 5.9. However, for larger lattice constants, the LRIGPW curve deviates slightly from the GPW results, suggesting increasing errors for longer C-C bond lengths.

To assess the accuracy of the LRI approach for atom pairs that are far apart, the dissociation curve for the HF monomer has been computed, see Figure 5.9(b). One could expect that at larger distances, when the interaction is still present, but the atoms are already separated, the fitting of pair densities becomes less accurate. Theoretically, numerical instabilities could lead to "density ripples" at distant points in space. In practice, such problems have not been observed. The GPW and LRIGPW dissociation curves agree perfectly. Also the dissociation limit, where the density between hydrogen and fluoride is quasi zero, is modeled accurately. The LRI basis sets are apparently flexible and diffuse enough to prevent a break down of LRI in the dissociation region. Stability issues have also never been encountered in the periodic case, which might be partly attributed to automatically set cutoffs for hardly overlapping

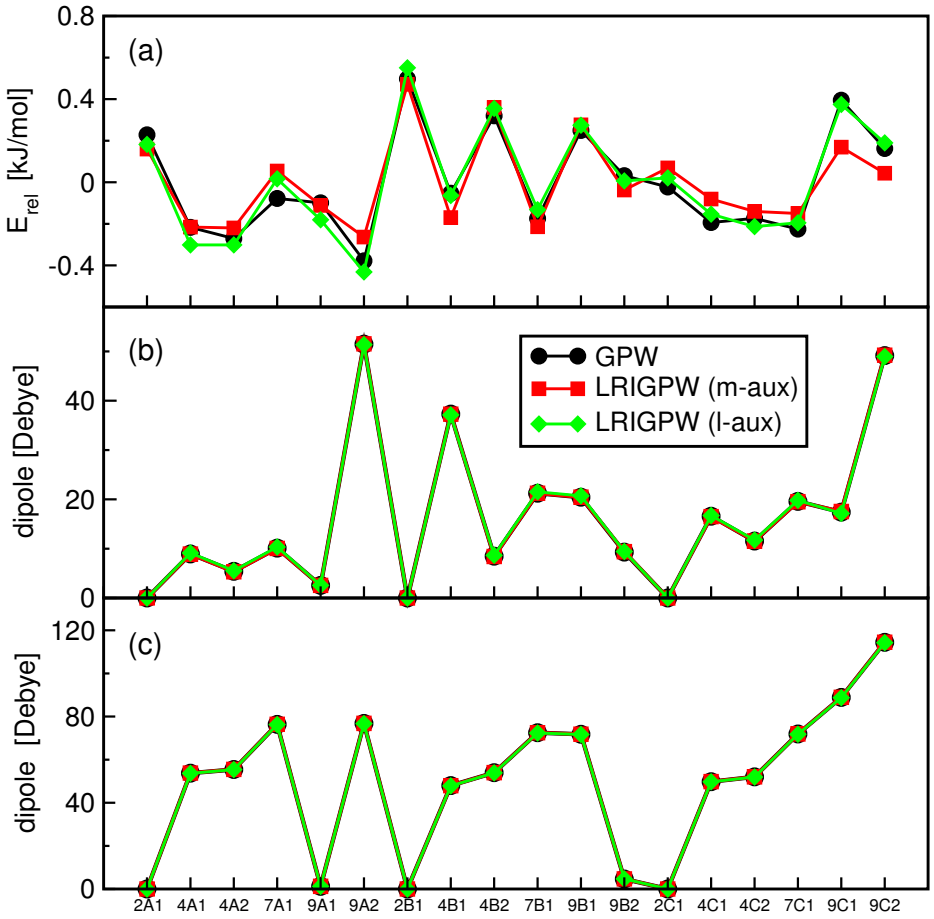

Figure 5.10.: Results for the 18 symmetry inequivalent ice XV structures. (a) Relative energies E_{rel} per molecule are defined as $(E - E_{\mathrm{aver}})/N_{\mathrm{mol}}$, where E is the total energy of the supercell, E_{aver} the average of all 18 structures and N_{mol} the number of molecules in the supercell. (b) Net dipole moments obtained by the Berry phase approximation. (c) Net dipole moments calculated from MLWO. Note that the curves in (b) and (c) are exactly on top of each other.

densities, i.e. pairs that are very far apart and that could establish a problem are not considered anyway.

5.5.4. Ice XV

The energetic differences between the 18 symmetry inequivalent ice XV structures are rather subtle and provide a tough test for the LRIGPW approach. In Figure 5.10(a) the total energies per molecule are reported with respect to the average. The black curve shown Figure 5.10(a) resembles closely the curve obtained with PBE in Ref. [224]. Small deviations must be attributed to different basis sets and the inclusion of the D3 correction. The differences between the 18 structures are in a small range of 1 kJ/mol. GPW and LRIGPW consistently predict that 9A2 is the

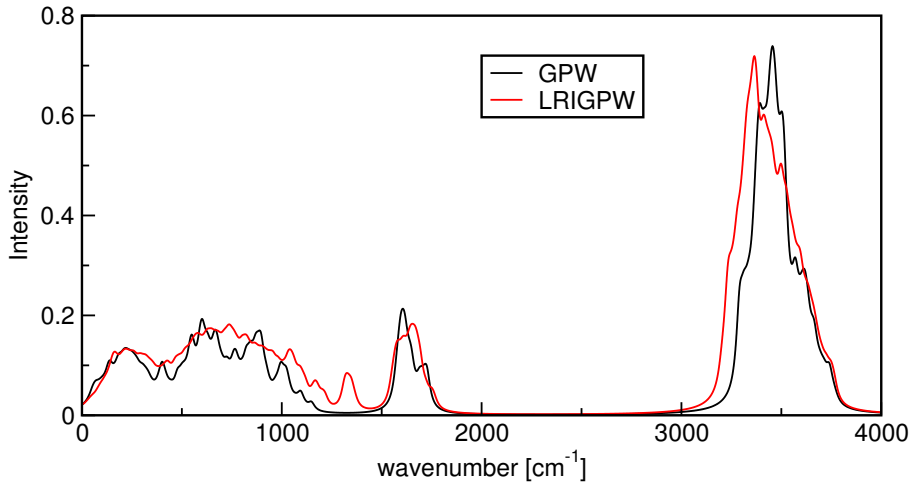

Figure 5.11.: IR-spectra obtained for the 7B1 structure of ice XV based on normal mode analysis. The spectra are broadened by Lorentzian functions with a full width at half maximum (FWHM) of 40cm^{-1}.

global minimum structure. Using the m-aux basis sets, deviations are observed for the structures with C-type and partly with A-type hydrogen bonding. Increasing the auxiliary basis sets, the accuracy is noticeably improved and only small deviations remain for the 4A1, 7A1 and 9A1 structures.

The net dipole moments obtained from the Berry phase approximation are presented in Figure 5.10(b). Their values seem to be insensitive to the inaccuracies introduced by the density fit. LRIGPW and GPW dipoles match exactly. In the periodic case, the Berry phase dipoles are defined modulo a certain period determined by the cell dimensions. In order to retrieve the unshifted values, the net dipole moments have been calculated from molecular dipoles estimated from MLWO, see Figure 5.10(c). Again, GPW and LRIGPW dipole moments agree perfectly. Furthermore, we find that 2A1, 2B1 and 2C1 have a net dipole moment of exactly zero, while 9A1 and 9B2 have very small residual dipoles.

The IR-spectrum of the 7B1 structure has been computed with GPW and LRIGPW and is displayed in Figure 5.11. The main spectral features are well reproduced by LRIGPW. However, the differences are noticeable, especially the additional peak at 1350 cm^{-1}. These inaccuracies are most likely attributed to the incremental approach used to obtain the spectra. The IR-spectra are obtained from static calculations within the harmonic approximation as outlined in Section 5.4. For such an approach, the derivatives of the forces (Hessian) and the derivatives of the dipole moments with respect to the atomic positions have to be computed. The intensities and shifts of the IR-bands are very sensitive to inaccuracies in these derivatives. In CP2K, the Hessian and the dipole derivatives are obtained from finite differences, which is less accurate than using an analytical approach. For LRIGPW, these derivatives are additionally affected by numerical noise when using the inverse of the overlap matrix $\mathbf{S}$. As dis-

cussed in Section 5.3, the overlap matrix has generally very large condition numbers for the auxiliary basis sets used, which limits the stability of the SCF convergence. In a finite difference approach, the atoms are incrementally shifted and the SCF is converged for each configuration. The differences in total energy between these increments can be in the range of the numerical accuracy when using simply the inverse of **S**. As a consequence, no distinct IR bands are obtained, but only numerical noise. However, the numerical stability is mostly recovered by computing the pseudoinverse for at least the most problematic atomic pairs. The numerical stability could be further improved by increasing the threshold for the accepted singular values when constructing the pseudoinverse. On the other hand, this would reduce the quality of the density fit as discussed in Section 5.3.3. When computing the derivatives with a finite difference approach, we carefully have to balance numerical stability and flexibility in the density fitting. This might have non-negligible effects on the IR-spectra which are very sensitive to errors in the Hessian and the dipole derivatives. However, a better agreement between LRIGPW and GPW spectra is expected if the derivatives would be calculated analytically.

5.5.5. Molecular crystals

Cohesive energies

The five investigated molecular crystals are characterized by intermolecular interactions of different nature, i.e., electrostatic polarization, hydrogen bonding, and dispersion forces. The computed and measured cohesive energies, E_{coh}, are listed in the upper part of Table 5.5. The experimental values are derived from sublimation energies ΔH_s. When comparing to experiment, it is to note that the temperature-dependence and zero-point vibrational energies are not considered for the computed values. The cohesive energies of molecular crystals have been calculated at the MP2 level previously and good agreement with experimental values is reported for the five crystals under investigation [54,227,230,231]. However, even at the PBE-D3 level, the computed E_{coh} compare reasonably well to experiment. In fact, the results for formic acid and succinic anhydride are in remarkably good agreement with the experimental values.

LRIGPW reproduces the cohesive energies computed with GPW very well. Absolute differences are smaller than 5 kJ/mol and relative deviations are smaller than 5%. By increasing the size of the auxiliary basis sets, the maximal deviation is reduced to less than 2 kJ/mol corresponding to a relative difference of 2%. The absolute errors are in any case within the margins of the CP correction E_{CP}, see Table 5.5. Note that E_{CP} is determined by the primary basis sets. Within the error range of the fit, the same E_{CP} are obtained with LRIGPW.

Table 5.5.: CP corrected cohesive energies E_{coh} (kJ/mol) for structures and cells relaxed at the GPW level. Lattice parameter *abc* (Å) of the unit cell for NH_3, CO_2, formic acid (FA), urea (U) and succinic anhydride (SA). The CP correction and the difference Δ_{GPW} with respect to GPW are given in kJ/mol as well. Note that for NH_3 and CO_2 $a = b = c$. δ_{GPW} is the ratio: difference/GPW value.

	GPW		LRIGPW (m-aux)		LRIGPW (l-aux)		Experiment*
	E_{coh}	CP	E_{coh}	Δ_{GPW}	E_{coh}	Δ_{GPW}	E_{coh}
NH_3	−44.3	−1.5	−44.9	0.6	−44.3	0.0	−36.3
CO_2	−24.1	−1.8	−23.9	0.2	−23.4	0.7	−31.1
FA	−68.7	−3.5	−70.8	2.1	−69.8	1.1	−68
U	−111.0	−4.2	−106.3	4.7	−111.0	0.0	−92
SA	−81.3	−3.5	−85.1	3.8	−83.2	1.9	−81
	GPW		LRIGPW (m-aux)		LRIGPW (l-aux)		Experiment†
	abc		abc	δ_{GPW}%	abc	δ_{GPW}%	abc
NH_3	5.035		4.941	1.9	4.986	1.0	5.048
CO_2	5.720		5.626	1.6	5.656	1.1	5.550
	10.328		–	–	10.236	0.9	10.241
FA	3.604		–	–	3.561	1.2	3.544
	5.374		–	–	5.354	0.4	5.356
	5.572		5.522	0.9	5.547	0.4	5.645
U	5.572		5.522	0.9	5.547	0.4	5.645
	4.704		4.704	0.0	4.692	0.2	4.704
	5.385		–	–	5.391	0.1	5.426
SA	6.960		–	–	6.954	0.1	6.975
	11.700		–	–	11.540	1.4	11.717

*for experimental E_{coh} see [227, 230] and SI of [231]; † for *abc* [226, 227, 230]

Lattice vectors

The cell relaxation with both, GPW and LRIGPW, preserves the orthorhombic or, in the case of NH_3 and CO_2, cubic symmetry. Only for urea, the lattice vectors deviate by 1-2° from an orthorhombic symmetry. Comparing the GPW results to experiment, relatively large deviations of up to 3% are observed. One source of error is for sure that we have neglected that the cell optimizations are also affected by the BSSE.

Using LRIGPW in combination with the m-aux basis sets, the lattice parameters did not convergence for formic acid and succinic anhydride and many steps have been required for the other crystals. Providing more flexibility by employing the l-aux basis

sets, the cell parameters converged smoothly and differ typically by less than 1% from the GPW results. The error introduced by the fit is probably not much larger than the BSSE. The latter might be severe since the cells have been relaxed at the DZVP level. Note that BSSE-free equilibrium lattice constants can be in principle obtained from an indirect approach, which is computationally less efficient, see Ref. [224, 227, 230] for details.

5.5.6. Computational efficiency of LRIGPW

The SCF convergence with GPW and LRIGPW requires approximately the same number of SCF steps to reach convergence. Slower convergence is only observed for systems with an extremely ill-conditioned overlap matrix $\mathbf{S}$, but the problem can be solved applying the pseudoinverse.

The total run times reported in Figure 5.12(a) refer to single point calculations of water boxes with N_{H_2O} molecules, which are carried out on $N_{H_2O}/2$ cores of a Cray XC40 machine. The major objective of LRIGPW is to reduce the computational cost for the grid-based operations as discussed in Section 5.2.1. And indeed, the time required for the collocation of the density and the integration of the potential is drastically reduced with LRIGPW and becomes in fact negligible, see Figure 5.12(a). The grid-operations are speeded up by a factor of 20-50 with respect to GPW. These speed-ups do not directly translate in an acceleration of the SCF iteration because LRI introduces also additional overhead. The density has to be fitted at each SCF step and Equation 5.9 has to be solved for each pair. Moreover, summing up the expansion coefficients in order to obtain $\{\tilde{a}_i^X\}$ requires additional MPI communication since the pairs are distributed.

The speed-up of the SCF step is system-dependent. For isolated systems, the speed-up is negligible since the number of pairs is not much larger than the number of atoms. On the contrary, a very large number of pairs is generated for periodic systems, like liquid water. For the medium-sized water boxes, the SCF step is accelerated by a factor of 3-5 depending on the platform, see Figure 5.12(b). The speed-up is noticeably smaller for large water boxes since the grid-operations contribute less to the total run time. Grid-operations require more than 90% of the total run time for 128 H_2O compared to 71% for 512 H_2O, see Figure 5.12(a).

The efficiency of LRIGPW depends also on the size of the auxiliary basis sets. Larger basis sets do not increase the time for the grid-operations significantly, but have an impact on the density fitting operations. The speed-up factors are a bit smaller for the l-aux basis set, but still in the range of 2-4, see Figure 5.12(b).

The acceleration of the SCF iterations is also to some extent platform-dependent which is due to the fact that the relative performance of particular operations depends on the system architecture. Namely, grid-operations on CrayXC40 cores are faster

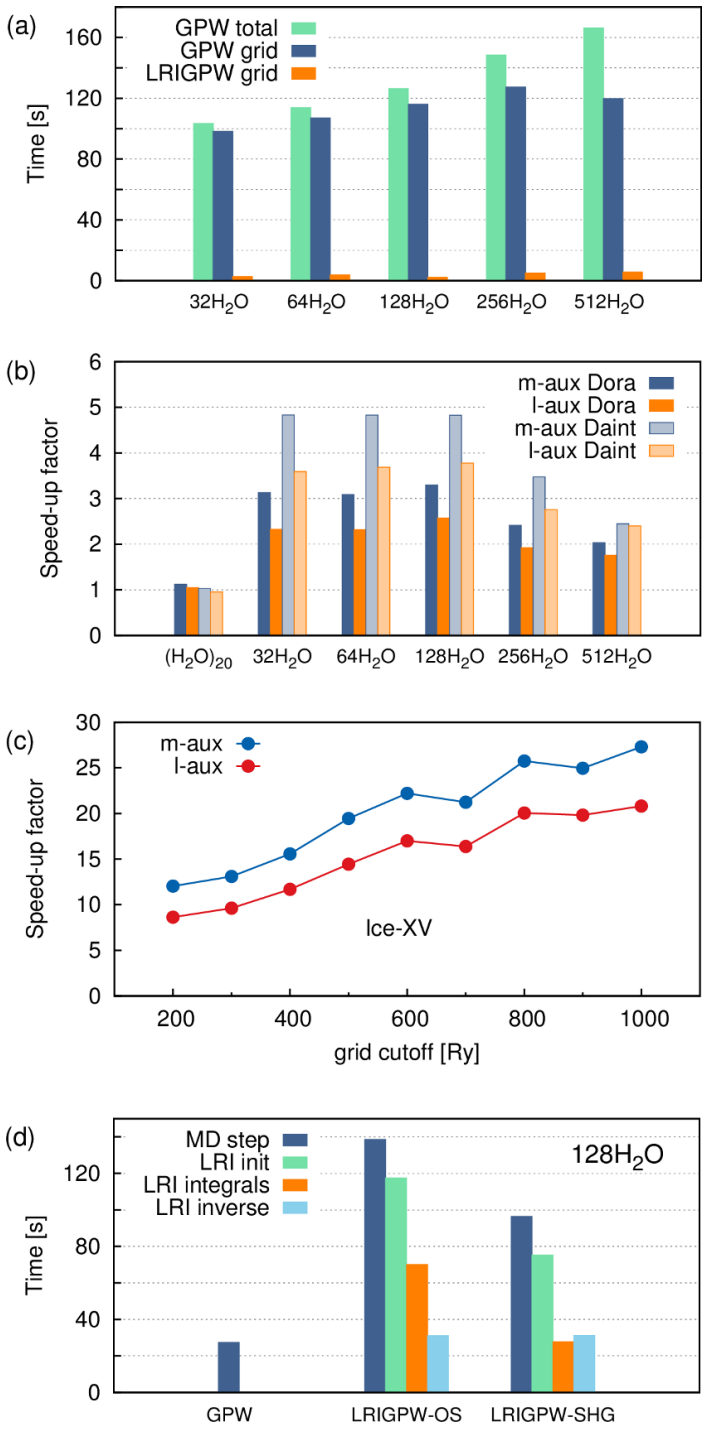

Figure 5.12.: Timings for GPW and LRIGPW on a Cray XC40 machine (Dora). (a) Total GPW run time and time for grid-based operations for liquid water boxes of different sizes. (b) Speed-up of the SCF step for an isolated water cluster of 20 molecules and liquid water on Cray XC30 (Daint) and Cray XC40 (Dora) machines. (c) Speed-up of the SCF step for ice XV dependent on the grid cutoff. (d) Time per MD step for liquid water (128 molecules) limiting the number of SCF steps to 5. *LRI init* indicates the time needed for initializing the LRI matrices prior to the SCF. For water boxes with N_{H_2O} molecules, $N_{H_2O}/2$ MPI processes are spanned. For $(H_2O)_{20}$ and ice XV, 8 and 32 processes are used, respectively.

relative to other operations, such as matrix multiplications etc., than on CrayXC30 cores. The possible speed-up we can gain with LRI is thus smaller.

In Figure 5.12(c) the SCF speed-up is reported dependent on the grid cutoff for the 2B1 structure of ice XV. Generally, the speed-ups are 5 times larger than for liquid water due to the triclinic symmetry of the ice XV boxes. For orthorhombic cells, grid-operations can be simplified by separating the spatial directions, thereby reducing the computational cost. The GPW grid-operations are thus more expensive for triclinic cells, while the computational cost is not significantly increased for the LRIGPW grid operations. Since the LRI overhead is independent of the cell symmetry, LRIGPW is more efficient with respect to GPW for non-orthorhombic boxes. LRIGPW becomes even more efficient when increasing the cutoff. For 1000 Ry, a factor of more than 25 is observed for the SCF speed-up.

The time per MD step for a liquid water box of 128 H_2O is reported in Figure 5.12(d). The number of SCF steps has been set to five to ensure a fair comparison. Even though the SCF is accelerated, the MD step takes 5 times longer with LRIGPW. This is due to the high computational cost related to the initialization of the LRI matrices. Almost 85% of the time is spent for calculating the LRI integrals and the inversion of the overlap matrix $\mathbf{S}$. The most time-consuming part is the calculation of $(a|b|\tilde{a})$, $(a|b|\tilde{b})$ and their derivatives. As mentioned before, an alternative integral scheme has been developed to replace the OS procedure. With the SHG scheme, the time for integrals and derivatives is already reduced by a factor of 2.5, see Figure 5.12(d). So far, the SHG implementation is merely a β-version and requires more optimization to be truly efficient. More effort is also required to minimize the computational cost for the inversion of $\mathbf{S}$.

The memory requirements of LRIGPW are also rather large since the precomputed LRI matrices are stored for each atom pair. This is not a major obstacle on high performance computing platforms, but on local computer clusters, and further optimization is necessary.

5.6. Conclusions

In this chapter, a local density fitting approach has been introduced within GPW in order to reduce the prefactor for the construction of the KS matrix. For periodic systems, the grid-based part of the calculation, such as the collocation of the density and integration of the potential, clearly dominates the calculation of the KS matrix and the overall run time. The computational demands for these operations are massively decreased with LRI since they are performed for each atom instead for each atom pair.

LRIGPW requires an independent set of auxiliary functions as input. Medium- and large-sized auxiliary basis sets have been generated from geometric progressions

without further need of optimization. Instability issues due to the size of these basis sets are addressed by computing the pseudoinverse instead of the inverse for ill-defined overlap matrices.

The accuracy of LRIGPW has been assessed for a wide range of systems. Total energies, reaction energies, intramolecular and intermolecular structure parameters are well reproduced by LRIGPW. This has also been found for the cohesive energies of five different molecular crystals. The difference to the GPW result is significantly smaller or at least not larger than the CP correction. Good agreement with respect to the GPW reference has been also found for the relative energies of the 18 symmetry inequivalent structures of ice XV. Especially, the net dipole moments are exactly reproduced. Furthermore, LRIGPW proved to be stable within the dissociation limit and applicable for lattice optimizations of 2D and 3D periodic structures. The m-aux basis sets appeared to be applicable for almost all systems and properties. Employing larger auxiliary basis sets improves the results, particularly for cell relaxations and structural properties extracted from MD trajectories such as RDFs. Generally, the m-aux basis sets yield accurate results for primary basis sets of DZVP as well as TZV2P quality. However, providing more flexibility by adding more functions might be occasionally necessary at the TZV2P level.

For LRIGPW, the computational cost of the grid-dependent operations becomes negligible. This is reflected in a speed-up of the SCF step which depends on the symmetry of the simulation cell, the size of the system, the grid cutoff and (slightly) on the computing platform. For orthorhombic cells, speed-up factors of 2-5 are observed for relatively low grid cutoffs, whereas the SCF is speeded-up by a factor of 15 for triclinic cells. However, the initialization of the LRI matrices prior to the SCF becomes especially time-consuming when forces are required. So far, the MD step is in fact slower with LRI. The optimization of the initialization is in progress and the first version of a new integral scheme showed already some success. Moreover, the SCF step can be further accelerated by improving the parallelization strategy and removing communication overhead. Further development of LRIGPW also includes the extension to meta-GGAs which requires some LRI specific modifications for the kinetic energy density τ.

6. Integral evaluation of contracted solid harmonic Gaussian functions

An integral scheme based on solid harmonic Gaussians is introduced for the efficient calculation of two-center overlap integrals and their derivatives. Such integrals are important for the local resolution of identity (LRI) approach discussed in Chapter 5. The evaluation of the overlap $(a|b)$ has been adapted from Ref. [232]. However, LRI requires also the computation of the integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, where two functions are centered at A and B, respectively. The mathematical expressions for the latter are derived based on the differentiation rules of the spherical tensor gradient operator and have been formally proved for arbitrary quantum numbers l. The scheme presented herein is superior to the Cartesian Gaussian-based Obara-Saika scheme. A speed-up of factor 2.5 is already observed for the integrals and their derivatives even though the current implementation is still inefficient and requires more optimization.

6.1. Introduction

Rapid evaluation of analytical integrals is important for a wide range of methods in quantum chemistry. The computation of overlap integrals is required for quantum mechanics (QM), molecular mechanics (MM) and hybrid QM/MM methods. Many *ab initio* codes use Gaussian functions as orbital basis and numerous integral schemes for the evaluation of Gaussian overlap integrals have been proposed [216, 232–241]. A popular approach is the Obara-Saika (OS) scheme [216] which employs a recursive formalism over Cartesian Gaussian functions. Recurrence relations are also used in the McMurchie-Davidson scheme [233] based on Hermite Gaussian functions. Commonly, QM codes utilize spherical harmonic Gaussians (SpHG) since for a given angular momentum l the number of Cartesian Gaussian functions is greater or equal to the number of SpHGs, i.e., $(l+1)(l+2)/2$ Cartesians compare to $2l+1$ spherical harmonics. Thus, intermediary integrals are computed in a primitive Cartesian Gaussian basis followed by a transformation to contracted SpHG integrals.

An alternative integral scheme employs contracted solid harmonic Gaussian (SHG) functions, which correspond to SpHG functions after appropriate manipulation of the normalization constant. The formulation of these integrals is based on the spherical tensor gradient operator (STGO) [242, 243] and Hobson's theorem of differentiation [244]. The resulting expressions contain an angular momentum part that is

independent of the exponent. This part is obtained by relatively simple recursions and can be pre-computed and re-used multiple times. The contraction is performed only for integrals of s-functions. Unlike for Cartesian functions, subsequent transformation and contraction steps are not required. The SHG-integral approach is superior to Cartesian-based methods especially for highly contracted orbital basis sets where the subsequent contraction step is otherwise dominant.

In this work, the SHG scheme is employed for the overlap integrals required for the LRIGPW (local resolution of identity for Gaussian and plane waves) method implemented in the CP2K program package [121,191], see Chapter 5. An SHG-based integral scheme is particularly efficient when LRIGPW is used in combination with the highly contracted Molopt basis sets [122]. These basis sets and the corresponding auxiliary basis sets are so-called family basis sets, i.e., the same set of exponents is used for all angular momentum states. In this case, the contraction vector can also be re-used multiple times. The evaluation of the two-center overlap integrals $(a|b)$ follows closely the procedure developed by Giese and York [232]. LRIGPW requires also overlap integrals of the type $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, where two different functions are located at center A and B, respectively. The extension of the expressions given by Giese *et al.* to these type of integrals is not straightforward and requires as intermediate step the evaluation of the integrals $(a|r_a^{2n}|b)$ and $(a|r_b^{2n}|b)$. The formulas for $(a|r_a^{2n}|b)$, $(a|r_b^{2n}|b)$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ are derived applying the STGO and are proved to be valid for all $n \in \mathbb{N}$. An implementation of a β-version of the SHG integrals and their derivatives is discussed. The performance of the SHG integrals is compared to the Cartesian-based OS scheme.

6.2. Integral and derivative evaluation

6.2.1. Definitions and notations

The notation used herein corresponds to Ref. [232,245,246]. The unnormalized, primitive SHG is given by

$$\chi_{l,m}(\mathbf{r},\alpha) = C_{l,m}(\mathbf{r}) \exp\left(-\alpha r^2\right) \tag{6.1}$$

where the complex solid harmonics $C_{l,m}(\mathbf{r})$,

$$C_{l,m}(\mathbf{r}) = \sqrt{\frac{4\pi}{2l+1}} r^l Y_{l,m}(\theta,\phi), \tag{6.2}$$

are obtained by rescaling the standard spherical harmonics $Y_{l,m}(\theta,\phi)$. The primitive SpHG functions $\phi_{l,m}$ are closely related to SHGs by

$$\phi_{l,m} = \sqrt{\frac{2l+1}{4\pi}} \chi_{l,m}. \tag{6.3}$$

Since the integrals in CP2K are in fact based on SpHGs, we have to correct for the factor $k = \sqrt{(2l+1)/4\pi}$ accordingly. Contracted SpHG functions $\varphi_{l,m}$ are constructed as linear combination of the primitive functions

$$\varphi_{l,m}(\mathbf{r}) = \tilde{N}_l \sum_i c_i \phi_{l,m}(\mathbf{r}, \alpha_i) = k\tilde{N}_l \sum_i c_i \chi_{l,m}(\mathbf{r}, \alpha_i), \tag{6.4}$$

where $\tilde{N}_l$ is the normalization constant given by

$$\tilde{N}_l = \left[\sum_i \sum_j \frac{\pi^{1/2}(2l+2)! c_i c_j}{2^{2l+3}(l+1)!(\alpha_i + \alpha_j)^{l+3/2}} \right]^{-1/2} \tag{6.5}$$

as defined in Ref. [247]. For convenience, the factor k is in the following included in the normalization constant N_l, i.e $N_l = k\tilde{N}_l$.

We further introduce the regular scaled solid harmonics $R_{l,m}$. The relationship between the regular solid harmonics $C_{l,m}$ and $R_{l,m}$ is

$$R_{l,m}(\mathbf{r}) = \frac{1}{\sqrt{(l-m)!(l+m)!}} C_{l,m}(\mathbf{r}). \tag{6.6}$$

The regular scaled solid harmonics are complex and can be decomposed into a real (cosine) and an imaginary (sine) part

$$R_{l,m} = R^c_{l,m}(\mathbf{r}) + iR^s_{l,m}(\mathbf{r}). \tag{6.7}$$

The cosine and sine parts can be constructed from recursion relations as discussed below. In the following, we use the notations,

$$\begin{aligned} \mathbf{r}_a &= \mathbf{r} - \mathbf{R}_a \\ \mathbf{r}_b &= \mathbf{r} - \mathbf{R}_b \\ r^2_{ab} &= |\mathbf{R}_a - \mathbf{R}_b|^2 \end{aligned} \tag{6.8}$$

where $\mathbf{R}_a$ references the position of the Gaussian center A and $\mathbf{R}_b$ the position of center B. The scalar derivative of $X(r^2_{ab})$ with respect to the square-separation r^2_{ab} between the two Gaussian centers is denoted by

$$X(r^2_{ab})^{(k)} = \left(\frac{\partial}{\partial r^2_{ab}} \right)^k X(r^2_{ab}). \tag{6.9}$$

6.2.2. Overlap integrals $(a|b)$

In this section, the expressions for the two-center overlap integrals of two contracted functions $\varphi_{l_a m_a}$ and $\varphi_{l_b m_b}$ centered at $\mathbf{R}_a$ and $\mathbf{R}_b$ are given. The derivation of the overlap integrals,

$$(a|b) = \int \varphi_{l_a m_a}(\mathbf{r}_a) \varphi_{l_b m_b}(\mathbf{r}_b) d\mathbf{r}, \tag{6.10}$$

follows the work by Giese and York [232] adapting changes related to different normalization constants accordingly.

Derivation of the integral expressions

The derivation is based on Hobson's theorem [244] of differentiation which states that

$$C_{l,m}(\nabla)f(r^2) = 2^l C_{l,m}(\mathbf{r})\left(\frac{\partial}{\partial r^2}\right)^l f(r^2), \tag{6.11}$$

where the differential operator $C_{l,m}(\nabla)$ is referred to as STGO. The derivation of the overlap integrals starts now by noting that $\exp(-\alpha r^2)$ is an eigenfunction of $C_{l,m}(\nabla)$ with the eigenvalue $(-2\alpha)^l$. Thus, the primitive SHG at center $\mathbf{R}_a$ can be rewritten as

$$\chi_{l,m}(\mathbf{r}_a) = \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l}, \tag{6.12}$$

where $C_{l,m}(\nabla_a)$ acts on $\mathbf{R}_a$. A solid harmonic s-function is simply given by $\chi_{0,0}(\mathbf{r}) = \exp\left(-\alpha r^2\right)$, which yields

$$\chi_{l,m}(\mathbf{r}_a) = \frac{C_{l,m}(\nabla_a)\chi_{0,0}(\mathbf{r}_a)}{(2\alpha)^l}. \tag{6.13}$$

Inserting this expression in Equation 6.10 gives the following formula for the overlap integrals

$$(a|b) = C_{l_a,m_a}(\nabla_a)C_{l_b,m_b}(\nabla_b)S(r_{ab}^2). \tag{6.14}$$

The integral $S(r_{ab}^2)$ corresponds to the contracted overlap of s-functions and depends on the square separation r_{ab}^2, see Equation 6.18. Considering this integral scheme as a Gaussian multipole expansion of a diffuse charge density, $S(r_{ab}^2)$ can be interpreted as the monopole result. The product and differentiation rules for the STGO [236–238, 248] are applied to Equation 6.14 and conversion to real expressions yields finally

$$\begin{aligned}(a|b) = (-1)^{l_b} A_{l_a m_a} A_{l_b m_b} \sum_{j=0}^{min(l_a,l_b)} & 2^{l_a+l_b-j} S(r_{ab}^2)^{(l_a+l_b-j)} \\ & \times (2j-1)!! \sum_{\kappa=0}^{j} B_{j,\kappa} Q_{l_a m_a, l_b m_b, j\kappa}^{c/s,c/s}(\mathbf{r}_{ab}),\end{aligned} \tag{6.15}$$

where the prefactors are

$$A_{l,m} = (-1)^m \sqrt{(2-\delta_{m,0})(l+m)!(l-m)!} \tag{6.16}$$

$$B_{j,\kappa} = \frac{1}{(2-\delta_{\kappa,0})(j+\kappa)!(j-\kappa)!}. \tag{6.17}$$

The scalar derivative of the monopole result is defined by

$$S(r_{ab}^2)^{(k)} = N_{l_a} N_{l_b} \sum_{\alpha\in \mathrm{A}} \sum_{\beta\in \mathrm{B}} \frac{c_\alpha c_\beta}{(2\alpha)^{l_a}(2\beta)^{l_b}} \left(\frac{\partial}{\partial r_{ab}^2}\right)^k (0_a|0_b), \tag{6.18}$$

denoting the overlap of two primitive s-functions as

$$(0_a|0_b) = \int \chi_{0,0}(\alpha, \mathbf{r}_a)\chi_{0,0}(\beta, \mathbf{r}_b)d\mathbf{r} = \left(\frac{\pi}{\alpha+\beta}\right)^{3/2} e^{-\zeta_{\alpha\beta}r_{ab}^2}, \tag{6.19}$$

where $\zeta_{\alpha\beta} = \alpha\beta/(\alpha+\beta)$. The expression for $(0_a|0_b)$ given in Equation 6.19 is used to initiate the following recurrence formula,

$$\left(\frac{\partial}{\partial r_{ab}^2}\right)^k (0_a|0_b) = (-\zeta_{\alpha\beta})\left(\frac{\partial}{\partial r_{ab}^2}\right)^{k-1} (0_a|0_b). \tag{6.20}$$

The notation $Q^{c/s,c/s}_{l_a m_a, l_b m_b, j\kappa}$ in Equation 6.15 refers to elements of the 2×2 matrix $\mathbf{Q}_{l_a m_a, l_b m_b, j\kappa}$ built from the real translation matrix described in the following part. Positive m reference a cosine (c) and negative m a sine (s) component which is discussed in detail below.

The expression for the overlap integrals can be easily generalized to integrals of the type

$$(a|\mathcal{O}|b) = \iint \varphi_{l_a m_a}(\mathbf{r}-\mathbf{R}_a)\mathcal{O}(\mathbf{r}-\mathbf{r}')\varphi_{l_b m_b}(\mathbf{r}'-\mathbf{R}_b)d\mathbf{r}d\mathbf{r}', \tag{6.21}$$

given that the integrals for $l_a, l_b = 0$ are a function of the square separation r_{ab}^2. Solely the expression for $S(r_{ab}^2)$ is modified replacing $(0_a|0_b)$ by $(0_a|\mathcal{O}|0_b)$. For more details see Ref. [232].

The real translation matrix

The matrix $\mathbf{Q}_{l_a m_a, l_b m_b, j\kappa}$ is obtained from the real translation matrix $\mathbf{W}_{lm,j\kappa}$ which depends solely on the angular momentum, but not on the exponent,

$$\mathbf{Q}_{l_a, m_a, l_b m_b, j\kappa} = \mathbf{W}_{l_a \mu_a, j\kappa}\mathbf{W}^{\mathrm{T}}_{l_b \mu_b, j\kappa}, \tag{6.22}$$

where the label μ indicates that we insert in this case only positive values,i.e. $\mu = |m|$. The real translation matrix is a 2×2 matrix with the elements

$$\mathbf{W}_{lm,j\kappa} = \begin{pmatrix} W^{cc}_{lm,j\kappa} & W^{cs}_{lm,j\kappa} \\ W^{sc}_{lm,j\kappa} & W^{ss}_{lm,j\kappa} \end{pmatrix} \tag{6.23}$$

The result of the matrix-matrix multiplication in Equation 6.22 is

$$\begin{aligned} \mathbf{Q}_{l_a, m_a, l_b m_b, j\kappa} &= \begin{pmatrix} W^{cc}_a W^{cc}_b + W^{cs}_a W^{cs}_b & W^{cc}_a W^{sc}_b + W^{cs}_a W^{ss}_b \\ W^{sc}_a W^{cc}_b + W^{ss}_a W^{cs}_b & W^{sc}_a W^{sc}_b + W^{ss}_a W^{ss}_b \end{pmatrix}_{(j\kappa)} \\ &= \begin{pmatrix} Q^{cc}_{ab,j\kappa} & Q^{cs}_{ab,j\kappa} \\ Q^{sc}_{ab,j\kappa} & Q^{ss}_{ab,j\kappa} \end{pmatrix} \end{aligned} \tag{6.24}$$

The resulting matrix contains four entries where positive m refer to a cosine (c) component and negative m to a sine (s) component, i.e.,

$$
\begin{aligned}
Q^{cc}_{l_a m_a, l_b m_b, j\kappa}&: \quad m_a, m_b \geq 0 \\
Q^{cs}_{l_a m_a, l_b m_b, j\kappa}&: \quad m_a > 0,\, m_b < 0 \\
Q^{sc}_{l_a m_a, l_b m_b, j\kappa}&: \quad m_a < 0,\, m_b > 0 \\
Q^{ss}_{l_a m_a, l_b m_b, j\kappa}&: \quad m_a < 0,\, m_b < 0.
\end{aligned}
$$

For example, the *cc* component is addressed for the overlap of $\varphi_{3,2}(\mathbf{r}_a)$ and $\varphi_{4,1}(\mathbf{r}_b)$ and *sc* for $\varphi_{3,-2}(\mathbf{r}_a)$ and $\varphi_{4,1}(\mathbf{r}_b)$.

The elements of the transformation matrix $\mathbf{W}_{lm,j\kappa}$ are also defined for negative m. The expressions for $\mathbf{W}_{lm,j\kappa}$ have been adapted from Ref. [245, 246] reversing the sign of $\mathbf{r}_{ab}$ for simplicity,

$$
\begin{aligned}
W^{cc}_{lm,j\kappa}(\mathbf{r}_{ab}) &= \left(\frac{1}{2}\right)^{\delta_{\kappa 0}} \left[R^c_{l-j,m-\kappa}(\mathbf{r}_{ab}) + (-1)^\kappa R^c_{l-j,m+\kappa}(\mathbf{r}_{ab})\right] \\
W^{cs}_{lm,j\kappa}(\mathbf{r}_{ab}) &= -R^s_{l-j,m-\kappa}(\mathbf{r}_{ab}) + (-1)^\kappa R^s_{l-j,m+\kappa}(\mathbf{r}_{ab}) \\
W^{sc}_{lm,j\kappa}(\mathbf{r}_{ab}) &= \left(\frac{1}{2}\right)^{\delta_{\kappa 0}} \left[R^s_{l-j,m-\kappa}(\mathbf{r}_{ab}) + (-1)^\kappa R^s_{l-j,m+\kappa}(\mathbf{r}_{ab})\right] \\
W^{ss}_{lm,j\kappa}(\mathbf{r}_{ab}) &= R^c_{l-j,m-\kappa}(\mathbf{r}_{ab}) - (-1)^\kappa R^c_{l-j,m+\kappa}(\mathbf{r}_{ab}).
\end{aligned}
\tag{6.25}
$$

The cosine $R^c_{l,m}(\mathbf{r})$ and sine component $R^s_{l,m}(\mathbf{r})$ of the regular scaled solid harmonics, see Equation 6.7, are obtained by relatively simple recursion relations [245, 246]

$$R^c_{00} = 1 \tag{6.26}$$

$$R^s_{00} = 0 \tag{6.27}$$

$$R^c_{l+1,l+1} = -\frac{xR^c_{ll} - yR^s_{ll}}{2l+2} \tag{6.28}$$

$$R^s_{l+1,l+1} = -\frac{yR^c_{ll} + xR^s_{ll}}{2l+2} \tag{6.29}$$

$$R^{c/s}_{l+1,m} = \frac{(2l+1)zR^{c/s}_{lm} - r^2 R^{c/s}_{l-1,m}}{(l+m+1)(l-m+1)}, \qquad 0 \leq m < l \tag{6.30}$$

where $\mathbf{r} = (x, y, z)$. The usage of c/s in the last recurrence formula indicates that the relation is used for both, $R^c_{lm}(\mathbf{r})$ and $R^s_{lm}(\mathbf{r})$. The recursions are only valid for positive m. However, the regular scaled solid harmonics are also defined for negative indices and satisfy the following symmetry relations

$$
\begin{aligned}
R^c_{l,-m} &= (-1)^m R^c_{lm} \\
R^s_{l,-m} &= -(-1)^m R^s_{lm}.
\end{aligned}
\tag{6.31}
$$

Note that these symmetry relations have to be employed for the evaluation of $R^{c/s}_{l-j,\mu-\kappa}$ since $\mu - \kappa$ can be also negative. Furthermore, only elements with $l - j \geq \mu \pm \kappa$ give non-zero contributions.

As is easily verified, the matrix elements of $\mathbf{W}_{lm,j\kappa}$ obey the same symmetry relations with respect to sign changes of m,

$$\begin{aligned} W^{cc}_{l\overline{m},jk} &= (-1)^m W^{cc}_{lm,jk} \\ W^{cs}_{l\overline{m},jk} &= (-1)^m W^{cs}_{lm,jk} \\ W^{sc}_{l\overline{m},jk} &= -(-1)^m W^{sc}_{lm,jk} \\ W^{ss}_{l\overline{m},jk} &= -(-1)^m W^{ss}_{lm,jk} \end{aligned} \tag{6.32}$$

where we have used the notation $\overline{m} = -m$. These symmetry relations are not important for the evaluation of $\mathbf{W}_{l\mu,j\kappa}$ since $\mu \geq 0$ by definition. However, they are required for the evaluation of the derivatives discussed below.

6.2.3. Integrals $(a|r_a^{2n}|b)$ and $(a|r_b^{2n}|b)$

The expressions for $(a|r_a^{2n}|b)$ and $(a|r_b^{2n}|b)$ are fundamental for the derivation of the formulas for $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ discussed in the next section. The computation of the integrals

$$(a|r_a^{2n}|b) = \int \varphi_{l_a m_a}(\mathbf{r}_a) r_a^{2n} \varphi_{l_b m_b}(\mathbf{r}_b) d\mathbf{r} \tag{6.33}$$

$$(a|r_b^{2n}|b) = \int \varphi_{l_a m_a}(\mathbf{r}_a) r_b^{2n} \varphi_{l_b m_b}(\mathbf{r}_b) d\mathbf{r}, \qquad n \in \mathbb{N} \tag{6.34}$$

is not straightforward. Equation 6.15 cannot be simply adapted by replacing $(0_a|0_b)$ with $(0_a|r_{a/b}^{2n}|0_b)$ since this would only be valid for operators of the type $\mathcal{O}(\mathbf{r} - \mathbf{r}')$.

The product $\chi_{l,m}(\mathbf{r}_a) r_a^{2n}$ in terms of the STGO

The operator r_a^{2n} depends on the position $\mathbf{R}_a$ and $C_{l,m}(\nabla_a)$ is acting on the product of $\chi_{l,m}(\mathbf{r}_a)$ and r_a^{2n},

$$\chi_{l,m}(\mathbf{r}_a) r_a^{2n} = C_{l,m}(\mathbf{r}_a) \exp(-\alpha \mathbf{r}_a) r_a^{2n}. \tag{6.35}$$

The expression of this product in terms of the STGO $C_{l,m}(\nabla_a)$ is non-trivial. The derivation is again based on the application of Hobson's theorem and has been attempted in the first place for specific n starting from n=0, see Equation 6.12. The solutions have been then built recursively up to n=5, see Appendix C.1. From these, the expression for generic n has been deduced,

$$\chi_{l,m}(\mathbf{r}_a) r_a^{2n} = \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l} \sum_{j=0}^{n} \binom{n}{j} \frac{(l+j-1)!}{(l-1)!\alpha^j} \exp\left(-\alpha r_a^2\right) r_a^{2(n-j)}, \tag{6.36}$$

and proved to be valid for all $n \in \mathbb{N}$, see Appendix C.2. Note that we obtain n terms for $\chi_{l,m}(\mathbf{r}_a) r_a^{2n}$ multiplied with the corresponding combinatorial factors.

Derivation of integral expressions

Inserting Equations 6.36 and 6.13 in Equation 6.33 yields

$$(a|r_a^{2n}|b) = C_{l_a,m_a}(\nabla_a)C_{l_b,m_b}(\nabla_b)T_a(r_{ab}^2), \tag{6.37}$$

where $T_a(r_{ab}^2)$ is again the monopole results for the given integral. The derivation follows now the same procedure as for the overlap $(a|b)$. The final result is

$$\begin{aligned}(a|r_a^{2n}|b) = (-1)^{l_b} A_{l_a m_a} A_{l_b m_b} \sum_{j=0}^{min(l_a,l_b)} & 2^{l_a+l_b-j} T_a(r_{ab}^2)^{(l_a+l_b-j)} \\ & \times (2j-1)!! \sum_{\kappa=0}^{j} B_{j,\kappa} Q^{c/s,c/s}_{l_a m_a, l_b m_b, j\kappa}(\mathbf{r}_{ab}),\end{aligned} \tag{6.38}$$

where the scalar derivative of $T_a(r_{ab}^2)$ is

$$T_a(r_{ab}^2)^{(k)} = N_{l_a} N_{l_b} \sum_{\alpha\in \mathrm{A}} \sum_{\beta\in \mathrm{B}} \frac{c_\alpha c_\beta}{(2\alpha)^{l_a}(2\beta)^{l_b}} \sum_{j=0}^{n} \binom{n}{j} \frac{(l_a+j-1)!}{(l_a-1)!\alpha^j} (0_a|r_a^{2(n-j)}|0_b)^{(k)}. \tag{6.39}$$

The integral over primitive s-functions is

$$\begin{aligned}(0_a|r_a^{2m}|0_b) &= \int \chi_{0,0}(\alpha, \mathbf{r}_a) r_a^{2m} \chi_{0,0}(\beta, \mathbf{r}_b) d\mathbf{r} \\ &= \frac{\pi^{3/2}\exp(-\zeta_{\alpha\beta} r_{ab}^2)}{2^m c^{m+3/2}} \sum_{j=0}^{m} 2^j \frac{(2m+1)!!\binom{m}{j}}{(2j+1)!!} \frac{\beta^{2j}}{c^j} r_{ab}^{2j}\end{aligned} \tag{6.40}$$

with $c = \alpha + \beta$ and $\zeta_{\alpha\beta} = \alpha\beta/c$. Equation 6.40 has been validated for $m = 0,...,5$; see Appendix C.3. A proof of the given formula for all $m \in \mathbb{N}$ is similarly elaborate as for Equation 6.36 and has not been accomplished yet. The derivatives of $(0_a|r_a^{2m}|0_b)$ are obtained by applying the Leibniz rule of differentiation [249]. The general formula and the explicit expressions for $0 \leq m \leq 5$ are given in Appendix C.3. Note that Equation 6.38 is identical to Equation 6.15 except that the integral $S(r_{ab}^2)$ is replaced by $T_a(r_{ab}^2)$ representing a sum of integrals. The formulas for $(a|r_b^{2n}|b)$ are analogous to $(a|r_a^{2n}|b)$ substituting $T_a(r_{ab}^2)$ by $T_b(r_{ab}^2)$,

$$\begin{aligned}(a|r_b^{2n}|b) = (-1)^{l_b} A_{l_a m_a} A_{l_b m_b} \sum_{j=0}^{min(l_a,l_b)} & 2^{l_a+l_b-j} T_b(r_{ab}^2)^{(l_a+l_b-j)} \\ & \times (2j-1)!! \sum_{\kappa=0}^{j} B_{j,\kappa} Q^{c/s,c/s}_{l_a m_a, l_b m_b, j\kappa}(\mathbf{r}_{ab}),\end{aligned} \tag{6.41}$$

where the derivative of $T_b(r_{ab}^2)$ with respect to r_{ab}^2 is

$$T_b(r_{ab}^2)^{(k)} = N_{l_a} N_{l_b} \sum_{\alpha\in \mathrm{A}} \sum_{\beta\in \mathrm{B}} \frac{c_\alpha c_\beta}{(2\alpha)^{l_a}(2\beta)^{l_b}} \sum_{j=0}^{n} \binom{n}{j} \frac{(l_b+j-1)!}{(l_b-1)!\beta^j} (0_a|r_b^{2(n-j)}|0_b)^{(k)}. \tag{6.42}$$

The expression for the integral over s-functions and the operator r_b^2 is given by

$$(0_a|r_b^{2m}|0_b) = \frac{\pi^{3/2}\exp(-\zeta_{\alpha\beta}r_{ab}^2)}{2^m c^{m+3/2}} \sum_{j=0}^{m} 2^j \frac{(2m+1)!!\binom{m}{j}}{(2j+1)!!} \frac{\alpha^{2j}}{c^j} r_{ab}^{2j}, \tag{6.43}$$

and has been as well only verified for $0 \leq m \leq 5$.

6.2.4. Overlap integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$

The two-center overlap integrals of the type $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ include two functions at the same center and are defined by

$$(a|b|\tilde{a}) = \int \varphi_{l_a m_a}(\mathbf{r}_a)\varphi_{\tilde{l}_a,\tilde{m}_a}(\mathbf{r}_a)\varphi_{l_b m_b}(\mathbf{r}_b)d\mathbf{r} \tag{6.44}$$

$$(a|b|\tilde{b}) = \int \varphi_{l_a m_a}(\mathbf{r}_a)\varphi_{l_b m_b}(\mathbf{r}_b)\varphi_{\tilde{l}_b,\tilde{m}_b}(\mathbf{r}_b)d\mathbf{r}. \tag{6.45}$$

The integrals $(a|b|\tilde{a})$ or $(a|b|\tilde{b})$ are not as easily derived for SHG functions as for Cartesian Gaussian functions. The complications arise from reducing the product of two functions located at the same atom to one function. For Cartesian Gaussians, the exponents and the angular momentum of the two functions are simply added as shown in Equation 5.42. However, this is not possible for SHGs. First, the expression of the product of two SHGs in terms of the STGO is derived. Then the integral expressions are rationalized.

Product of two SHGs at the same center

The derivation for two primitive SHG functions at $\mathbf{R}_a$ starts by using the definitions given in Equations 6.1 and 6.2,

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)\chi_{\tilde{l},\tilde{m}}(\tilde{\alpha},\mathbf{r}_a) &= C_{l,m}(\mathbf{r}_a)\exp(-\alpha r_a^2)C_{\tilde{l},\tilde{m}}(\mathbf{r}_a)\exp(-\tilde{\alpha}r_a^2)\\ &= \lambda\exp\left(-\alpha' r_a^2\right)Y_{l,m}Y_{\tilde{l},\tilde{m}}r_a^{l+\tilde{l}}\end{aligned} \tag{6.46}$$

where $\alpha' = \alpha + \tilde{\alpha}$ and

$$\lambda = \sqrt{\frac{4\pi}{2l+1}}\sqrt{\frac{4\pi}{2\tilde{l}+1}}. \tag{6.47}$$

The product of two spherical harmonics can be expanded in terms of spherical harmonics,

$$Y_{l,m}Y_{\tilde{l},\tilde{m}} = \sum_{L,M} G_{M,m,\tilde{m}}^{L,l,\tilde{l}} Y_{L,M}, \tag{6.48}$$

where $G_{M,m,\tilde{m}}^{L,l,\tilde{l}}$ are the Clebsch-Gordon (CG) coefficients and $|l-\tilde{l}| \leq L \leq l+\tilde{l}$. Note that this expansion of the product of spherical harmonics is valid since they form a

complete set of orthonormal functions. A similar expansion for solid harmonics is not possible since this would require the completeness of the latter in $\mathbb{R}^3$. Inserting the CG expansion in Equation 6.46 (1), re-introducing solid harmonics (2) as defined in 6.2 and employing definition 6.1 (3) yields

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)\chi_{\tilde{l},\tilde{m}}(\tilde{\alpha},\mathbf{r}_a) &\overset{(1)}{=} \lambda\exp\left(-\alpha' r_a^2\right)\sum_{L,M} G^{L,l,\tilde{l}}_{M,m,\tilde{m}} Y_{L,M} r_a^{l+\tilde{l}} \\ &\overset{(2)}{=} \lambda\sum_{L,M} G^{L,l,\tilde{l}}_{M,m,\tilde{m}}\sqrt{\frac{2L+1}{4\pi}} C_{L,M}(\mathbf{r}_a)\exp\left(-\alpha' r_a^2\right) r_a^{l+\tilde{l}-L} \\ &\overset{(3)}{=} \lambda\sum_{L,M} G^{L,l,\tilde{l}}_{M,m,\tilde{m}}\sqrt{\frac{2L+1}{4\pi}} \chi_{L,M}(\alpha',\mathbf{r}_a) r_a^{l+\tilde{l}-L}.\end{aligned} \tag{6.49}$$

The L quantum numbers of the non-zero contributions in the CG expansion proceed in steps of two starting from $L_{\max} = l + \tilde{l}$. Thus, $l + \tilde{l} - L$ is always even and we can express $\chi_{L,M}(\alpha', \mathbf{r}_a) r_a^{l+\tilde{l}-L}$ in terms of the STGO using Equation 6.36,

$$\begin{aligned}&\chi_{l,m}(\alpha,\mathbf{r_a})\chi_{\tilde{l},\tilde{m}}(\tilde{\alpha},\mathbf{r_a}) \\ &\quad= \lambda\sum_{L,M} G^{L,l,\tilde{l}}_{M,m,\tilde{m}}\sqrt{\frac{2L+1}{4\pi}}\frac{C_{L,M}(\nabla_a)}{(2\alpha')^L}\sum_{j=0}^{p}\binom{p}{j}\frac{(L+j-1)!}{(L-1)!(\alpha')^j}\exp\left(-\alpha' r_a^2\right) r_a^{2(p-j)}\end{aligned} \tag{6.50}$$

employing the notation $p = (l + \tilde{l} - L)/2$.

Derivation of integral expressions

Having an expression for the product of primitive SHG functions, the further steps in the derivation of the $(a|b|\tilde{a})$ integrals are analogous to the $(a|b)$ integrals. Inserting the STGO formulations given in Equation 6.50 and Equation 6.12 in Equation 6.44 yields

$$(a|b|\tilde{a}) = \sum_{L_a,M_a} G^{L_a,l_a,\tilde{l}_a}_{M_a,m_a,\tilde{m}_a} C_{L_a,M_a}(\nabla_a) C_{l_b,m_b}(\nabla_b) P_a(L_a, r_{ab}^2) \tag{6.51}$$

with

$$\begin{aligned}P_a(L_a, r_{ab}^2) =& \lambda\sqrt{\frac{2L_a+1}{4\pi}} N_{l_a} N_{l_b} N_{\tilde{l}_a}\sum_{\alpha\in\mathrm{A}}\sum_{\beta\in\mathrm{B}}\sum_{\tilde{\alpha}\in\mathrm{A}}\frac{c_\alpha c_\beta c_{\tilde{\alpha}}}{(2\alpha')^{L_a}(2\beta)^{l_b}} \\ &\times\sum_{j=0}^{p}\binom{p}{j}\frac{(L_a+j-1)!}{(L_a-1)!(\alpha')^j}(0_{a'}|r_a^{2(p-j)}|0_b),\end{aligned} \tag{6.52}$$

where the expression for integrals over primitive s-functions,

$$(0_{a'}|r_a^{2m}|0_b) = \int \chi_{0,0}(\alpha',\mathbf{r}_a) r_a^{2m} \chi_{0,0}(\beta,\mathbf{r}_b) d\mathbf{r}, \tag{6.53}$$

is given in Equation 6.40. The derivation proceeds as for the $(a|b)$ and $(a|r_a^{2n}|b)$ integrals yielding the final formula,

$$\begin{aligned}(a|b|\tilde{a}) = (-1)^{l_b} A_{l_b m_b} \sum_{L_a,M_a} G^{L_a,l_a,\tilde{l}_a}_{M_a,m_a,\tilde{m}_a} A_{L_a M_a} \sum_{j=0}^{min(L_a,l_b)} 2^{L_a+l_b-j} \\ \times P_a(L_a, r_{ab}^2)^{(L_a+l_b-j)} (2j-1)!! \sum_{\kappa=0}^{j} B_{j,\kappa} Q^{c/s,c/s}_{L_a M_a, l_b m_b, j\kappa}(\mathbf{r}_{ab}).\end{aligned} \tag{6.54}$$

The integral $(a|b|\tilde{a})$ can be considered as a sum of $(a|r_a^{2n}|b)$ integrals, introducing some modifications due to normalization and contraction. The expression for $(a|b|\tilde{b})$ is analogous and is obtained by replacing P_a by P_b, defined in Equation 6.56, and performing the CG expansion for the two functions centered at $\mathbf{R}_b$, which yields

$$\begin{aligned}(a|b|\tilde{b}) = A_{l_a m_a} \sum_{L_b,M_b} G^{L_b,l_b,\tilde{l}_b}_{M_b,m_b,\tilde{m}_b} (-1)^{L_b} A_{L_b M_b} \sum_{j=0}^{min(l_a,L_b)} 2^{l_a+L_b-j} \\ \times P_b(L_b, r_{ab}^2)^{(l_a+L_b-j)} (2j-1)!! \sum_{\kappa=0}^{j} B_{j,\kappa} Q^{c/s,c/s}_{l_a m_a, L_b M_b, j\kappa}(\mathbf{r}_{ab})\end{aligned} \tag{6.55}$$

with

$$\begin{aligned}P_b(L_b, r_{ab}^2) =& \lambda \sqrt{\frac{2L_b+1}{4\pi}} N_{l_a} N_{l_b} N_{\tilde{l}_b} \sum_{\alpha\in \mathrm{A}} \sum_{\beta\in \mathrm{B}} \sum_{\tilde{\beta}\in \mathrm{B}} \frac{c_\alpha c_\beta c_{\tilde{\beta}}}{(2\alpha)^{l_a} (2\beta')^{L_b}} \\ & \times \sum_{j=0}^{p} \binom{p}{j} \frac{(L_b+j-1)!}{(L_b-1)!(\beta')^j} (0_a|r_b^{2(p-j)}|0_{b'})\end{aligned} \tag{6.56}$$

where $\beta' = \beta + \tilde{\beta}$.

6.2.5. Derivatives

The derivatives of the integrals $(a|b)$, $(a|r_a^{2n}|b)$, $(a|r_b^{2n}|b)$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ are obtained by applying the product rule to the r_{ab}^2-dependent quantity and the angular-dependent part. The derivatives for $(a|b)$ with respect to $\mathbf{R}_a$ are

$$\begin{aligned}\frac{\partial}{\partial R_{a,i}}(a|b) =& 2(R_{a,i} - R_{b,i}) \sum_{j=0}^{min(l_a,l_b)} S(r_{ab}^2)^{(l_a+l_b-j+1)} \widetilde{Q}^{c/s,c/s}_{l_a m_a, l_b m_b}(j) \\ & + \sum_{j=0}^{min(l_a,l_b)} S(r_{ab}^2)^{(l_a+l_b-j)} \frac{\partial \widetilde{Q}^{c/s,c/s}_{l_a m_a, l_b m_b}(j)}{\partial R_{a,i}}\end{aligned} \tag{6.57}$$

with $i = x, y, z$ and where we have introduced the notation

$$\widetilde{Q}^{c/s,c/s}_{l_a m_a, l_b m_b}(j) = (-1)^{l_b} A_{l_a m_a} A_{l_b m_b} 2^{l_a+l_b-j} (2j-1)!! \sum_{\kappa=0}^{j} B_{j,\kappa} Q^{c/s,c/s}_{l_a m_a, l_b m_b, j\kappa}. \tag{6.58}$$

The derivatives of $(a|r_a^{2n}|b)$ and $(a|r_b^{2n}|b)$ are obtained from Equations 6.57 by substituting $S(r_{ab}^2)$ by $T_a(r_{ab}^2)$ and $T_b(r_{ab}^2)$, respectively. For $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, we replace by $P_a(L_a, r_{ab}^2)$ and $P_b(L_b, r_{ab}^2)$ considering additionally the CG expansion.

For the derivatives of $\widetilde{Q}^{c/s,c/s}_{l_a m_a, l_b m_b, j}$, the differentiation rules of the real scaled solid harmonics $R_{l,m}$, see Ref. [250], are employed,

$$\begin{aligned}
\frac{\partial}{\partial x} R^{c/s}_{l,m} &= \frac{1}{2}\left(R^{c/s}_{l-1,m+1} - R^{c/s}_{l-1,m-1}\right) \\
\frac{\partial}{\partial y} R^{c/s}_{l,m} &= \frac{(\pm 1)_m}{2}\left(R^{s/c}_{l-1,m+1} + R^{s/c}_{l-1,m-1}\right) \\
\frac{\partial}{\partial z} R^{c/s}_{l,m} &= R^{c/s}_{l-1,m},
\end{aligned} \tag{6.59}$$

where $(\pm 1)_m = 1$ if $m \geq 0$ and $(\pm 1)_m = -1$ if $m < 0$. Note that the derivative of $R^c_{l,m}$ with respect to y is constructed from the sine and $R^s_{l,m}$ from the cosine components. Using the differentiation rules for $R_{l,m}$, the Cartesian derivatives of $\widetilde{Q}^{c/s,c/s}_{l_a m_a, l_b m_b, j}$ can be expressed in terms of lower angular momentum functions,

$$\begin{aligned}
\frac{\partial \widetilde{Q}^{c/s,c/s}_{l_a,m_a,l_b,m_b}(j)}{\partial x_a} = \quad & \frac{A_{l_a,m_a}}{A_{l_a-1,m_a+1}} \widetilde{Q}^{c/s,c/s}_{l_a-1,m_a+1,l_b,m_b}(j) \\
& - \frac{A_{l_a,m_a}}{A_{l_a-1,m_a-1}} \widetilde{Q}^{c/s,c/s}_{l_a-1,m_a-1,l_b,m_b}(j) \\
& - \frac{A_{l_b,m_b}}{A_{l_b-1,m_b+1}} \widetilde{Q}^{c/s,c/s}_{l_a,m_a,l_b-1,m_b+1}(j) \\
& + \frac{A_{l_b,m_b}}{A_{l_b-1,m_b-1}} \widetilde{Q}^{c/s,c/s}_{l_a,m_a,l_b-1,m_b-1}(j)
\end{aligned} \tag{6.60}$$

$$\begin{aligned}
\frac{\partial \widetilde{Q}^{c/s,c/s}_{l_a,m_a,l_b,m_b}(j)}{\partial y_a} = \quad & (\pm 1)_{m_a} \frac{A_{l_a,m_a}}{A_{l_a-1,m_a+1}} \widetilde{Q}^{\mathbf{s/c},c/s}_{l_a-1,m_a+1,l_b,m_b}(j) \\
& + (\pm 1)_{m_a} \frac{A_{l_a,m_a}}{A_{l_a-1,m_a-1}} \widetilde{Q}^{\mathbf{s/c},c/s}_{l_a-1,m_a-1,l_b,m_b}(j) \\
& - (\pm 1)_{m_b} \frac{A_{l_b,m_b}}{A_{l_b-1,m_b+1}} \widetilde{Q}^{c/s,\mathbf{s/c}}_{l_a,m_a,l_b-1,m_b+1}(j) \\
& - (\pm 1)_{m_b} \frac{A_{l_b,m_b}}{A_{l_b-1,m_b-1}} \widetilde{Q}^{c/s,\mathbf{s/c}}_{l_a,m_a,l_b-1,m_b-1}(j)
\end{aligned} \tag{6.61}$$

$$\begin{aligned}
\frac{\partial \widetilde{Q}^{c/s,c/s}_{l_a,m_a,l_b,m_b}(j)}{\partial z_a} = \quad & 2\frac{A_{l_a,m_a}}{A_{l_a-1,m_a}} \widetilde{Q}^{c/s,c/s}_{l_a-1,m_a,l_b,m_b}(j) \\
& - 2\frac{A_{l_b,m_b}}{A_{l_b-1,m_b}} \widetilde{Q}^{c/s,c/s}_{l_a,m_a,l_b-1,m_b}(j).
\end{aligned} \tag{6.62}$$

A special case has to be considered for the x, y components, when $m = 0$. The matrices of the type $Q^{c/s,c/s}_{l_a,-1,l_b,m_b,j\kappa}$ or $Q^{c/s,c/s}_{l_a,m_a,l_b,-1,j\kappa}$ are never calculated since the real translation matrix $\mathbf{W}_{l\mu,j\kappa}$ is evaluated only for positive m, see Equation 6.22. However, we can use the symmetry relations given in Equations 6.32. For example if $m_a = 0$, the following relations are used for the x-derivative

$$\begin{aligned} \widetilde{Q}^{cc}_{l_a,-1,l_b,m_b} &= (-1)\widetilde{Q}^{cc}_{l_a,1,l_b,m_b} \\ \widetilde{Q}^{cs}_{l_a,-1,l_b,m_b} &= (-1)\widetilde{Q}^{cs}_{l_a,1,l_b,m_b} \end{aligned} \tag{6.63}$$

and for the y derivative we employ the symmetry relations:

$$\begin{aligned} \widetilde{Q}^{sc}_{l_a,-1,l_b,m_b} &= -(-1)\widetilde{Q}^{sc}_{l_a,1,l_b,m_b} \\ \widetilde{Q}^{ss}_{l_a,-1,l_b,m_b} &= -(-1)\widetilde{Q}^{ss}_{l_a,1,l_b,m_b}. \end{aligned} \tag{6.64}$$

Note that calculating the Cartesian derivatives of the SHG integrals requires only to increase the scalar derivative of the monopole integral by +1 and to combine the lower-order matrix elements of $Q^{c/s,c/s}_{l_a m_a,l_b m_b,j\kappa}$ linearly. Compared to Cartesian-based integral schemes, the effort for the recursive part of the calculation is not increased.

6.3. Implementation details

The SHG integrals have been implemented within the LRIGPW framework described in Chapter 5. For LRIGPW, the integrals $(a|b)$, $(\tilde{a}|\tilde{b})$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ are required where a, b reference primary basis functions and $\tilde{a}, \tilde{b}$ auxiliary basis functions.

In order to obtain $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, the two SHG functions at the same center are summarized employing the CG expansion given in Equation 6.48. The CG coefficients for SHGs are computed by multiplying Equation 6.48 by $Y_{L,M}(\theta,\phi)$ and integrating

$$G^{L,l,\tilde{l}}_{M,m,\tilde{m}} = \int Y_{L,M}(\theta,\phi) Y_{l,m}(\theta,\phi) Y_{\tilde{l},\tilde{m}}(\theta,\phi) d\Omega, \tag{6.65}$$

where θ and ϕ are the angular coordinates of the spherical polar system. Note that we assume in Equation 6.48 an expansion in real spherical functions and insert actually real rather than complex spherical harmonics in Equation 6.65. The permissible values for L range in steps of 2 from $|l - \tilde{l}|$ to $l + \tilde{l}$, but not all possible combinations with M give non-zero contributions. For $l, \tilde{l} \leq 2$, the product of two spherical harmonics is expanded in not more than four terms. However, the number of terms progresses with increasing $l + \tilde{l}$.

The pseudocode for the implementation of the SHG integrals for LRIGPW is shown in Figure 6.1. Since the CG coefficients do not depend on the atom positions, the non-vanishing CG coefficients are pre-calculated before entering the loop over pairs and are re-used for each atom pair. The matrix $\widetilde{\mathbf{Q}}_{l_1 m_1,l_2 m_2}(j, \mathbf{r}_{ab})$, defined in Equation 6.58, and its Cartesian derivatives are only calculated once per

Tabulate CG coefficients $G^{L,l,\tilde{l}}_{M,m,\tilde{m}}$

Loop over all pairs AB

$l_{\max} = \mathrm{MAX}(l_{a,\max} + \tilde{l}_{a,\max}, l_{b,\max} + \tilde{l}_{b,\max})$

Pre-tabulate $R^c_{l,m}(\mathbf{r}_{ab})$ and $R^s_{l,m}(\mathbf{r}_{ab}) \quad 0 \leq l \leq l_{\max}$

Pre-calculate matrix $\widetilde{\mathbf{Q}}_{l_1 m_1, l_2 m_2}(j, \mathbf{r}_{ab}) \quad 0 \leq (l_1, l_2) \leq l_{\max}$

Compute derivatives of $S(r^2_{ab})$ and $\tilde{S}(r^2_{ab})$

Compute derivatives of $P_a(r^2_{ab})$ and $P_b(r^2_{ab})$

Calculate $(a|b) = \sum_j \left\{S_a(r^2_{ab})\right\}' \widetilde{Q}^{c/s,c/s}_{l_a m_a, l_b m_b}(j, \mathbf{r}_{ab})$

Calculate $(\tilde{a}|\tilde{b}) = \sum_j \left\{\tilde{S}_a(r^2_{ab})\right\}' \widetilde{Q}^{c/s,c/s}_{\tilde{l}_a \tilde{m}_a, \tilde{l}_b \tilde{m}_b}(j, \mathbf{r}_{ab})$

Calculate $(a|b|\tilde{a}) = \sum_{L_a, M_a} G^{L_a, l_a, \tilde{l}_a}_{M_a, m_a, \tilde{m}_a} \sum_j \left\{P_a(L_a, r^2_{ab})\right\}' \widetilde{Q}^{c/s,c/s}_{L_a M_a, l_b m_b}(j, \mathbf{r}_{ab})$

Calculate $(a|b|\tilde{b}) = \sum_{L_b, M_b} G^{L_b, l_b, \tilde{l}_b}_{M_b, m_b, \tilde{m}_b} \sum_j \left\{P_b(L_b, r^2_{ab})\right\}' \widetilde{Q}^{c/s,c/s}_{l_a m_a, L_b M_b}(j, \mathbf{r}_{ab})$

End AB loop

Figure 6.1.: Pseudocode for calculating SHG integrals $(a|b)$, $(\tilde{a}|\tilde{b})$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ for LRIGPW, where a, b refer to functions of the primary and $\tilde{a}, \tilde{b}$ to functions of the auxiliary basis set set. The notation $\{X(r^2_{ab})\}'$ indicates that the $(l_1{+}l_2{-}j)$-th derivative of X with respect to r^2_{ab} is taken.

pair because they are independent of the exponents and can be used for $(a|b)$, $(\tilde{a}|\tilde{b})$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$. Then, the scalar derivatives of the integrals over primitive s-functions are computed and contracted yielding $S(r^2_{ab})^{(k)}$, $\tilde{S}(r^2_{ab})^{(k)}$, $P_a(r^2_{ab})^{(k)}$ and $P_b(r^2_{ab})^{(k)}$. The integrals and their Cartesian derivatives are finally computed from the contracted s-integrals and $\widetilde{\mathbf{Q}}_{l_1 m_1, l_2 m_2}(j, \mathbf{r}_{ab})$. The evaluation of $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ is so far restricted to $l_{a/b} + \tilde{l}_{a/b} = 11$ because the calculation of $P_{a/b}(r^2_{ab})$ is limited to $p = (l_{a/b} + \tilde{l}_{a/b} - L_{a/b})/2 \leq 5$, see Equations 6.52 and 6.56. In practice, this is more than sufficient.

Potentially, the SHG scheme is superior to the Cartesian-based OS scheme, especially for highly contracted basis sets. Within the OS scheme, the intermediate integrals of primitive Cartesian Gaussian functions are computed and subsequently contracted and transferred to integrals of SpHGs. For SHG integrals on the other hand, the contraction is only performed for the integrals over s-functions and the transformation to SpHGs is trivially accomplished by manipulating the normalization constant. Furthermore, the Cartesian integrals are computed recursively for each primitive, whereas in the SHG scheme recursive operations are only required for evaluating $R^{c/s}_{l,m}$. The computational cost for computing the latter is negligible since they can be pre-tabulated and re-used multiple times. The SHG scheme gets especially efficient when derivatives of the integrals are required. The derivatives of Cartesian Gaussians are constructed from higher-order angular functions (i.e., $l + 1$) whereas the derivatives of $R^{c/s}_{l,m}$ are constructed from combinations of $(l - 1)$ harmonics. In

Table 6.1.: Ratio of the integral timings comparing the Obara-Saika (OS) and the solid harmonic Gaussian (SHG) scheme for different primary and auxiliary basis sets. *m-aux* indicates the medium-sized and *l-aux* the large auxiliary basis set. The time for evaluating all integrals, i.e. $(a|b)$, $(\tilde{a}|\tilde{b})$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, and their derivatives has been measured.

ratio	DZVP		TZVP		TZV2P	
	m-aux	l-aux	m-aux	l-aux	m-aux	l-aux
$\mathrm{OS}_{(\text{screened})}/\mathrm{SHG}$	2.4	2.4	1.6	1.7	1.1	1.1
$\mathrm{OS}_{(\text{not screened})}/\mathrm{SHG}$	5.0	5.1	3.4	3.4	2.2	2.2

the OS scheme, the transfer to contracted spherical functions has to be performed for each spatial direction. The contraction step is a bottleneck for highly-contracted basis sets since it is 4 times more expensive when also derivatives are computed. The same contracted quantities, i.e., the scalar derivatives of $S(r_{ab}^2)$ and $P_{a/b}(r_{ab}^2)$, are used for the integrals as well as their Cartesian derivatives and almost no additional effort is necessary.

So far, the implementation of the SHG integrals is not very efficient concerning array accesses which can be improved by more favorable indexing. Further optimization is also necessary for the contraction loops by employing, among others, a proper screening, i.e., neglecting contributions from primitives with negligible overlap.

6.4. Results and discussion

The time spent for evaluating the integrals $(a|b)$ and $(\tilde{a}|\tilde{b})$ is minimal compared to the computational cost for the integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$, independent of the integral scheme. Within the SHG scheme, the time required for computing matrix $\widetilde{\mathbf{Q}}_{l_1m_1,l_2m_2}(j,\mathbf{r}_{ab})$, see Equation 6.58, and its derivatives is negligible. Contracting the integrals over s-functions is the dominating step followed by the construction of the integrals from $P_{a/b}(r_{ab}^2)$ and $\widetilde{\mathbf{Q}}_{l_1m_1,l_2m_2}(j,\mathbf{r}_{ab})$.

The performance of the SHG scheme has been compared with the commonly used OS scheme. The tests have been performed for a box of 128 water molecules with the setup and basis sets described in Section 5.4. The time for evaluating all integrals $(a|b)$, $(\tilde{a}|\tilde{b})$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ and their derivatives is averaged over 10 molecular dynamics steps. The ratio between OS and SHG timings are reported for different primary and auxiliary basis sets in Table 6.1. For the OS scheme, contributions of hardly overlapping primitive Gaussians are screened by default. Such an approximation is not applied for the SHG method, yet. The screening speeds up the OS integral evaluation by a factor of two. Nevertheless, the SHG scheme is already faster by a factor of 2.4 at the DZVP level. For a fair comparison, the screening of the OS

integrals has been switched off showing that the SHG scheme is five times faster at the same level of accuracy. The performance of the SHG scheme with respect to the OS method is only slightly affected by the size of the auxiliary basis sets. However, the speed-up gained by the SHG scheme is much smaller with larger primary basis sets. The TZVP and TZV2P Molopt basis set contain the same number of exponents and the maximal angular momentum is unchanged. However, TZVP and TZV2P include more contractions. Most of the time is spent for computing $P_a(r_{ab}^2)$ and $P_b(r_{ab}^2)$, i.e., for contracting the integrals over primitive s-functions. This is due to the fact that the contraction loops are inefficiently implemented and employ no screening. With a more efficient implementation of this step, the reverse trend is expected, i.e., increasing speed-up with increasing number of contractions.

All results presented here have to be considered preliminary. Speed-ups are so far only obtained when calculating the integrals *and* their derivatives. No speed-up is observed for the integrals alone. This is not unexpected since the SHG scheme is especially efficient when derivatives are needed for the reasons discussed above.

6.5. Conclusions

An integral scheme based on SHG functions as been developed for the efficient evaluation of overlap integrals $(a|b)$, $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ and their derivatives important for the LRIGPW scheme introduced in Chapter 5. The expressions for $(a|b)$ have been adapted from the work of Giese and York [232]. Pursuing the application of the STGO in Ref. [232], formulas for the integrals $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ have been derived. It has been shown that the latter can be considered as sum of $(a|r_a^{2n}|b)$ and $(a|r_b^{2n}|b)$ integrals, where the operators r_a^{2n} and r_b^{2n} depend on the positions of centers A and B. The integral formulas for $(a|r_a^{2n}|b)$ and $(a|r_b^{2n}|b)$ have been developed by employing Hobson's theorem and have been proved to be valid for all $n \in \mathbb{N}$.

The SHG method has been shown, by measure of empirical timings, to be superior to the Cartesian Gaussian-based OS scheme for non-trivial angular momentums. Even though the implementation of the SHG scheme is still rather inefficient, speed-ups of a factor of 2.5 have been achieved for the LRI integrals and their derivatives. Using SHG functions leads to integral and derivative expressions where the angular part is independent of the primitive exponents. The angular part can thus be pre-computed and re-used multiple times. In addition, the calculation of the integrals and their derivatives requires only the contraction of the integrals over s-functions and their scalar derivatives.

7. Summary and outlook

7.1. Summary

The common focus of the work presented in this thesis is the introduction and application of methods aiming at the reduction of the complexity of the charge density within density functional theory (DFT). This has been accomplished with two different strategies. One approach is to divide the chemical model in two subsystems, where one subsystem is treated at the quantum mechanics (QM) level and the other at the molecular mechanics (MM) level. Such hybrid approaches combining different levels of theory are commonly referred to as QM/MM. The second strategy is to employ a resolution of identity (RI) approach to approximate the total charge density.

Modeling the electrostatic interactions between QM and MM subsystem has remained a challenge over the last decades. The first part of this work has been dedicated to the development of an image charge (IC) augmented QM/MM approach that explicitly includes polarization effects in the simulation of adsorbate-metal systems. The adsorbate is thereby described by DFT and the metallic substrate at the MM level of theory. The polarization effects at the metallic interface are accounted for by applying a self-consistent image charge formalism. The IC-QM/MM approach has been successfully employed to study organic and inorganic molecules adsorbed on Au(111) and Pt(111) surfaces. The method has been also applied to study the liquid water/Pt(111) interface by means of molecular dynamics (MD). Simulations at timescales larger than 100 ps can be addressed with IC-QM/MM, which would not be feasible with *ab initio* MD. Thus, IC-QM/MM is a valuable tool to study electrochemical interfaces where dynamical processes are typically occurring in the nanosecond range.

A similar QM/MM scheme has been employed to study the wetting of water on hexagonal boron nitride (h-BN) monolayers on Rh(111). The water film is treated at the DFT level and the substrate is described by MM. Induction effects play a minor role for the interactions between water and h-BN/Rh(111). A standard electrostatic embedding QM/MM approach is sufficient in this case. The electrostatic potential of the substrate has been reproduced by assigning partial charges to each substrate atom. To obtain such charges, a restrained electrostatic potential (RESP) fitting technique has been developed for periodic systems. The periodic RESP fitting is based on the Gaussian and plane waves (GPW) approach and offers enough flexibility to easily adjust charges for nano-structured substrates. The RESP charges have been

determined for two different interaction modes between h-BN and the metal, which are characterized by different modulations of the electrostatic potential over the surface. These charges have been used to perform MD simulations and to investigate effects at the interface between liquid water film and substrate. The local variations of the electrostatic potential at the surface were found to determine changes in the hydrophilicity of the substrate and to affect the surface tension and thus the contact angle.

The last part of this work pursued the second strategy for reducing the complexity of the DFT charge density. A local resolution of identity (LRI) approach has been introduced within the GPW framework. The dual representation of the DFT charge density in GPW provides an efficient way to linearize the scaling of the Kohn-Sham (KS) matrix. Adding the LRI approach on top of GPW, the linear scaling behavior is maintained, but the prefactor for expensive operations is reduced. In particular, the collocation of the density as well as the integration of the potential on the real-space grid dominate the entire computational effort in GPW for large periodic systems. The LRI implementation has been proposed to reduce the cost of the real-space grid operations, making this part of the computation inexpensive. LRIGPW is especially efficient for periodic structures with speed-ups of the self-consistent field (SCF) iterations by factors between 2-25. The speed-up depends mostly on the symmetry of the simulation cell, the number of grid points and the system size. In the present implementation, LRIGPW is affected by an excessive overhead in the initialization of the LRI integrals and especially their derivatives. For this reason, its application for MD simulations is still not efficient. In order to solve these problems, a new scheme for the analytic integration has been developed, which is based on solid harmonic Gaussian functions. The new scheme proved already to be superior to the initially implemented Cartesian Gaussian-based method. However, further optimizations of the new integration procedures and of other parts of the code are required before LRIGPW can be considered a broadly applicable alternative to the standard GPW scheme. Including LRI in GPW comes without significant loss of accuracy. The error introduced by the fit is usually smaller than the error due to the incompleteness of the basis sets.

In conclusion, both strategies, QM/MM and LRI, have been introduced to enabled the simulation of large systems. The system sizes that can be addressed with QM/MM will still be out of range with full DFT, whether LRI is used or not. However, QM/MM trades accuracy for speed and involves partly rather crude approximations, limiting the validity of the model. An accurate description of the electronic structure is inevitable for an adequate investigation of chemical problems. In this perspective, the development of more efficient full QM approaches, e.g., by means of RI approaches, remains important.

7.2. Outlook

Several algorithmic improvements of the methods developed in this thesis are possible and partly in progress. IC-QM/MM as implemented here is up to a factor of two slower than standard QM/MM. A more efficient procedure to evaluate the IC matrix is presently under development. The computational cost for the IC correction becomes negligible with the new scheme which is especially appealing from a user's point of view. As indicated before, also the LRIGPW approach is subject to further optimization.

The LRIGPW approach represents also a good basis for possible extensions. In particular, local density fitting is also applicable within the augmented Gaussian and plane waves (GAPW) framework [251, 252] developed for all-electron calculations. The GAPW approach is based on a partition of the density in smooth contributions and rapidly varying terms close to the nuclei. The construction of the GAPW density contributions relies on projections of basis functions. Such projections can be simplified using the LRI approach, thereby reducing the computational cost.

A. Derivatives for the Siepmann-Sprik potential

The last term of the pair potential v_2, see Equation 3.13, depends on the orientation of the dipole vector of the water molecule and is in the following referred to as d_2. The d_2 term and the three-body term v_3, see Equation 3.16, and their derivatives had to be explicitly implemented in CP2K. The assignment of the different vectors is given in Figure A.1.

A.1. Derivatives of the d_2 term

The d_2 term depends on the angle ϕ_{ji} between the dipole vector of the water molecule and vector $\mathbf{r}_{ji}$ and is defined as

$$d_2 = -Df_2(r_{ji})r_{ji}^{-3}\Phi_{ji}(\phi_{ji}), \tag{A.1}$$

where

$$f_2(r_{ji}) = \exp\left[\frac{B}{r_{ji} - r_{cut}}\right] \tag{A.2}$$

and

$$\Phi_{ji}(\phi_{ji}) = \exp\left[-8\left(\frac{cos\phi_{ji} - 1}{4}\right)^4\right] \tag{A.3}$$

The direction of the dipole vector is approximated by the vector $\mathbf{r}_{ix}$ lying in the H_2O molecular plane intersecting the HOH angle,

$$\mathbf{r}_{ix} = \mathbf{r}_{im} + \mathbf{r}_{il} \tag{A.4}$$

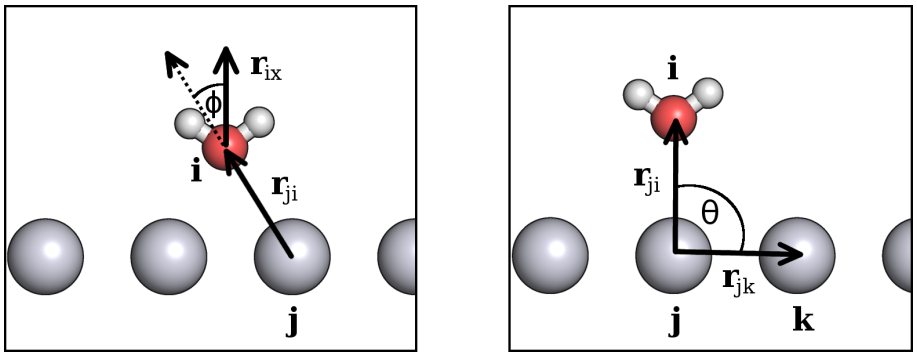

Figure A.1.: Definition of the vectors and angles for the Siepmann-Sprik potential.

The indices m and l refer to the two hydrogen atoms of the water molecule. Using the product rule, the derivative of d_2 with respect to i (O) and j (Pt) contains three terms. For the derivative with respect to m, l (H atoms) we only get one term,

$$\begin{aligned}\frac{\partial d_2}{\partial \mathbf{r}_j(n)} &= \tilde{A}_{j,n} + \tilde{B}_{j,n} + \tilde{C}_{j,n} \\ \frac{\partial d_2}{\partial \mathbf{r}_i(n)} &= \tilde{A}_{i,n} + \tilde{B}_{i,n} + \tilde{C}_{i,n} \\ \frac{\partial d_2}{\partial \mathbf{r}_{m/l}(n)} &= \tilde{C}_{m/l,n}.\end{aligned} \tag{A.5}$$

where $n = x, y, z$. The first term is

$$\begin{aligned}\tilde{A}_{j,n} &= -D \cdot \frac{\partial f_2(r_{ji})}{\partial r_{ji}} \cdot r_{ji}^{-3} \Phi_{ji}(\phi_{ji}) \cdot \frac{-\mathbf{r}_{ji}(n)}{r_{ji}} \\ \tilde{A}_{i,n} &= -\tilde{A}_{j,n},\end{aligned} \tag{A.6}$$

where

$$\frac{\partial f_2(r_{ji})}{\partial r_{ji}} = \frac{-B \exp\left[\frac{B}{r_{ji}-r_{cut}}\right]}{(r_{ji}-r_{cut})^2}. \tag{A.7}$$

The second term reads

$$\begin{aligned}\tilde{B}_{j,n} &= 3D \cdot f_2(r_{ji}) \cdot r_{ji}^{-4} \Phi_{ji}(\phi_{ji}) \cdot \frac{-\mathbf{r}_{ji}(n)}{r_{ji}} \\ \tilde{B}_{i,n} &= -\tilde{B}_{j,n}.\end{aligned} \tag{A.8}$$

The third term, $\tilde{C}$, includes the derivative of the function Φ_{ji}, which depends on $\mathbf{r}_i$, $\mathbf{r}_j$ and the positions of the hydrogen atoms $\mathbf{r}_m$ and $\mathbf{r}_l$.

$$\tilde{C}_{(j,i,m/l),n} = -D \cdot f_2(r_{ji}) \cdot r_{ji}^{-3} \cdot \Phi'_{ij}(\phi_{ij}), \tag{A.9}$$

where the derivative of Φ_{ji} is indicated by Φ'_{ji},

$$\Phi'_{ji}(\phi_{ji}) = \Phi_{ji}(\phi_{ji}) \cdot (-8) \cdot \left(\frac{\cos\phi_{ji} - 1}{4}\right)^3 \cdot (\cos\phi_{ji})'. \tag{A.10}$$

The cosine is defined as

$$\cos\phi_{ji} = \frac{\mathbf{r}_{ji}\mathbf{r}_{ix}}{r_{ji}r_{ix}} \tag{A.11}$$

and its derivative with respect to the hydrogen atoms m and l is,

$$\frac{\partial \cos\phi_{ji}}{\partial \mathbf{r}_{m/l}(n)} = \frac{\mathbf{r}_{ji}(n)}{r_{ji}r_{ix}} - \cos\phi_{ji}\frac{\mathbf{r}_{ix}(n)}{r_{ix}^2}. \tag{A.12}$$

Note that we obtain identical derivatives for both hydrogen atoms. The derivatives with respect to i and j are

$$\frac{\partial \cos\phi_{ji}}{\partial \mathbf{r}_j(n)} = \frac{-\mathbf{r}_{ix}(n)}{r_{ji}r_{ix}} + \cos\phi_{ji}\frac{\mathbf{r}_{ji}(n)}{r_{ji}^2} \tag{A.13}$$

$$\frac{\partial \cos\phi_{ji}}{\partial \mathbf{r}_i(n)} = -\frac{\partial \cos\phi_{ji}}{\partial \mathbf{r}_j(n)} - 2\frac{\partial \cos\phi_{ji}}{\partial \mathbf{r}_m(n)}. \tag{A.14}$$

A.2. Derivatives of the v_3 term

The three-body term is given by

$$v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = Ef_2(r_{ji})r_{ji}^{-\beta}\exp\left[F\cos^2\left(\frac{\theta_{ijk}}{2}\right)\right] \tag{A.15}$$

with the angle

$$\theta_{ijk} = \arccos\left(\frac{\mathbf{r}_{ji}\mathbf{r}_{jk}}{r_{ji}r_{jk}}\right). \tag{A.16}$$

The derivative with respect to Pt atom j and O atom i is a sum of three terms. The derivative with respect to Pt atom k contains only one term,

$$\begin{aligned}
\frac{\partial v_3}{\partial \mathbf{r}_j(n)} &= \tilde{E}_{j,n} + \tilde{F}_{j,n} + \tilde{G}_{j,n} \\
\frac{\partial v_3}{\partial \mathbf{r}_i(n)} &= \tilde{E}_{i,n} + \tilde{F}_{i,n} + \tilde{G}_{i,n} \\
\frac{\partial v_3}{\partial \mathbf{r}_k(n)} &= \tilde{G}_{k,n}.
\end{aligned} \tag{A.17}$$

The first term, $\tilde{E}$, is

$$\begin{aligned}
\tilde{E}_{j,n} &= E\cdot\frac{\partial f_2(r_{ji})}{\partial r_{ji}}\cdot r_{ji}^{-\beta}\cdot\exp\left[F\cos^2\left(\frac{\theta_{ijk}}{2}\right)\right]\cdot\frac{-\mathbf{r}_{ji}(n)}{r_{ji}} \\
\tilde{E}_{i,n} &= -\tilde{E}_{j,n}.
\end{aligned} \tag{A.18}$$

The second term, $\tilde{F}$, contains the derivative of $r_{ji}^{-\beta}$,

$$\begin{aligned}
\tilde{F}_{j,n} &= E\cdot f_2(r_{ji})\cdot(-\beta)r_{ji}^{(-\beta-1)}\cdot\exp\left[F\cos^2\left(\frac{\theta_{ijk}}{2}\right)\right]\cdot\frac{-\mathbf{r}_{ji}(n)}{r_{ji}} \\
\tilde{F}_{i,n} &= -\tilde{F}_{j,n}.
\end{aligned} \tag{A.19}$$

The third term, $\tilde{G}$, includes the derivative of angle θ_{ijk}, which depends on the positions of i, j and k,

$$\tilde{G}_{(j,i,k),n} = -2F\cdot v_3\cdot\cos\left(\frac{\theta_{ijk}}{2}\right)\cdot\sin\left(\frac{\theta_{ijk}}{2}\right)\cdot(\theta'_{ijk}/2), \tag{A.20}$$

where the derivative of θ_{ijk} is

$$(\theta'_{ijk}/2) = \left(\frac{1}{2}\arccos\left(\frac{\mathbf{r}_{ji}\mathbf{r}_{jk}}{r_{ji}r_{jk}}\right)\right)' = -\frac{1}{2\sqrt{1-(\cos\theta_{ijk})^2}}\cdot(\cos\theta_{ijk})'. \tag{A.21}$$

The derivatives of the cosine term are

$$\frac{\partial\cos\theta_{ijk}}{\partial\mathbf{r}_i(n)} = \frac{\mathbf{r}_{jk}(n)}{r_{ji}r_{jk}} - \cos\theta_{ijk}\frac{\mathbf{r}_{ji}(n)}{r_{ji}^2} \tag{A.22}$$

$$\frac{\partial\cos\theta_{ijk}}{\partial\mathbf{r}_k(n)} = \frac{\mathbf{r}_{ji}(n)}{r_{ji}r_{jk}} - \cos\theta_{ijk}\frac{\mathbf{r}_{jk}(n)}{r_{jk}^2} \tag{A.23}$$

$$\frac{\partial\cos\theta_{ijk}}{\partial\mathbf{r}_j(n)} = -\left(\frac{\partial\cos\theta_{ijk}}{\partial\mathbf{r}_i(n)} + \frac{\partial\cos\theta_{ijk}}{\partial\mathbf{r}_k(n)}\right). \tag{A.24}$$

B. Supplementary information for LRIGPW

B.1. Derivation of the fit formulas

Including the charge constraint in Equation 5.7 yields the modified minimization,

$$0 = \frac{d}{da_x}\left[\int(\rho_{\mathrm{AB}} - \tilde{\rho}_{\mathrm{AB}})^2 d\mathbf{r} + \lambda\left(\int \tilde{\rho}_{\mathrm{AB}} d\mathbf{r} - N_{\mathrm{AB}}\right)\right] \tag{B.1}$$

Taking the derivatives with respect to the expansion coefficients a_x gives

$$0 = -2\int(\rho_{\mathrm{AB}} - \tilde{\rho}_{\mathrm{AB}})\frac{d\tilde{\rho}_{\mathrm{AB}}}{da_x} d\mathbf{r} + \lambda\int\frac{d\tilde{\rho}_{\mathrm{AB}}}{da_x} d\mathbf{r} \tag{B.2}$$

with

$$\frac{d\tilde{\rho}_{\mathrm{AB}}}{da_x} = f_x \tag{B.3}$$

and

$$\int\frac{d\tilde{\rho}_{\mathrm{AB}}}{da_x} d\mathbf{r} = n_x. \tag{B.4}$$

Note that the derivative has to be taken with respect to a_i^{A} and a_j^{B} and that d/da_x is used in a generic fashion here. Inserting Equation B.3 and Equation B.4, the minimization condition reads

$$\begin{aligned} 0 &= 2\int\rho_{\mathrm{AB}} f_x d\mathbf{r} - 2\int\tilde{\rho}_{\mathrm{AB}} f_x d\mathbf{r} - \lambda n_x \\ &= 2\sum_{\mu\in\mathrm{A}\nu\in\mathrm{B}} P_{\mu\nu}\int\chi_\mu^A\chi_\nu^B f_x d\mathbf{r} - 2\sum_i a_i\int f_i f_x d\mathbf{r} - \lambda n_x \\ &= 2\sum_{\mu\in\mathrm{A}\nu\in\mathrm{B}} P_{\mu\nu}T_{\mu\nu x} - 2\sum_i a_i\tilde{S}_{ix} - \lambda n_x. \end{aligned} \tag{B.5}$$

Rearranging yields a linear system of equations,

$$2\sum_i S_{xi}a_i = 2t_x - \lambda n_x. \tag{B.6}$$

where S_{xi}, t_x and n_x are as defined in Section 5.2.2. In order to find an expression for the Lagrangian multiplier λ, Equation B.6 is multiplied from the left with the inverse

of $\mathbf{S}$, i.e. with $\sum_x S_{jx}^{-1}$,

$$\begin{aligned}\lambda \sum_x S_{jx}^{-1} n_x &= -2 \sum_i \underbrace{\sum_x S_{jx}^{-1} S_{xi}}_{E_{ji}} a_i + 2 \sum_x S_{jx}^{-1} t_x \\ &= -2a_j + 2 \sum_x S_{jx}^{-1} t_x,\end{aligned} \tag{B.7}$$

where E_{ji} indicates the identity matrix. Further multiplication with $\sum_j n_j$ gives

$$\lambda \sum_j \sum_x n_j S_{jx}^{-1} n_x = -2 \sum_j n_j a_j + 2 \sum_j \sum_x n_j S_{jx}^{-1} t_x. \tag{B.8}$$

After a final rearrangement we get

$$-\frac{\lambda}{2} \left(\mathbf{n}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{n} \right) = N_{\mathrm{AB}} - \mathbf{n}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{t}, \tag{B.9}$$

which is equivalent to

$$\tilde{\lambda} = \frac{N_{\mathrm{AB}} - \mathbf{n}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{t}}{\mathbf{n}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{n}}. \tag{B.10}$$

Table B.1.: Number of s, p, d, f and g functions in the auxiliary basis sets tailored for the primary basis sets of the Molopt type.

element	m-aux	m^+-aux	l-aux
H	10s9p8d6f	10s9p8d7f6g	10s10p10d7f
B	10s9p8d7f6g	–	–
C	15s13p12d11f9g	–	15s15p15d15f15g
N	15s13p12d11f9g	–	15s15p15d15f15g
O	15s13p12d11f9g	–	15s15p15d15f15g
F	15s13p11d10f8g	–	15s15p15d15f15g
Cl	15s13p11d10f8g	–	15s15p15d15f15g
Br	12s11p10d9f8g	–	12s12p12d12f12g

B.2. Auxiliary basis sets

The LRI auxiliary basis sets are tailored for the Molopt basis sets and given in CP2K format. Basis sets with the ending "-MEDIUM", "MEDIUM_PLUS" and "-LARGE" are abbreviated with m-aux, m^+-aux and l-aux, respectively. The size of the auxiliary basis sets is summarized in Table B.1. The CP2K basis set format is:

```
Element symbol  Name of the basis set
nset (repeat the following block of lines nset times)
dummy lmin lmax nexp nshell(lmin) nshell(lmin+1) ... nshell(lmax-1) nshell(lmax)
a(1)      c(1,l,1)      c(1,l,2) ...      c(1,l,nshell(l)), l=lmin...lmax
a(2)      c(2,l,1)      c(2,l,2) ...      c(2,l,nshell(l)), l=lmin...lmax
 .          .             .                 .                       .
 .          .             .                 .                       .
 .          .             .                 .                       .
a(nexp-1) c(nexp-1,l,1) c(nexp-1,l,2) ... c(nexp-1,l,nshell(l)), l=lmin...lmax
a(nexp)   c(nexp,l,1)   c(nexp,l,2)   ... c(nexp,l,nshell(l))  , l=lmin...lmax


nset     : Number of exponent sets
dummy    : dummy argument for historical reasons, no meaning
lmax     : Maximum angular momentum quantum number l
lmin     : Minimum angular momentum quantum number l
nshell(l): Number of shells for angular momentum quantum number l
a        : Exponent
c        : Contraction coefficient
```

The auxiliary basis sets are uncontracted and contain thus only one exponent per set and all contraction coefficients are set to one. Taking for example the third set of the m-aux basis set of element H, we have an s, p and d function, each with an exponent of 5.698118029748.

Hydrogen

```
H  LRI-MOLOPT-GTH-MEDIUM
10
2   0   0   1  1
  22.956000679816  1.0
2   0   1   1  1  1
  11.437045132575  1.0  1.0
2   0   2   1  1  1  1
   5.698118029748  1.0  1.0  1.0
2   0   2   1  1  1  1
   2.838893149810  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   1.414381779030  1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.704667527549  1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.351076584656  1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.174911945670  1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.087143916955  1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.043416487268  1.0  1.0  1.0  1.0
```

```
H  LRI-MOLOPT-GTH-LARGE
10
2   0   2   1  1  1  1
  22.956000679816   1.0  1.0  1.0
2   0   2   1  1  1  1
  11.437045132575   1.0  1.0  1.0
2   0   2   1  1  1  1
   5.698118029748   1.0  1.0  1.0
2   0   3   1  1  1  1  1
   2.838893149810   1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   1.414381779030   1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.704667527549   1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.351076584656   1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.174911945670   1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.087143916955   1.0  1.0  1.0  1.0
2   0   3   1  1  1  1  1
   0.043416487268   1.0  1.0  1.0  1.0
```

```
H  LRI-MOLOPT-GTH-MEDIUM-PLUS
10
2   0   0   1  1
  22.956000679816   1.0
2   0   1   1  1  1
  11.437045132575   1.0   1.0
2   0   2   1  1  1  1
   5.698118029748   1.0   1.0   1.0
2   0   3   1  1  1  1  1
   2.838893149810   1.0   1.0   1.0   1.0
2   0   4   1  1  1  1  1  1
   1.414381779030   1.0   1.0   1.0   1.0   1.0
2   0   4   1  1  1  1  1  1
   0.704667527549   1.0   1.0   1.0   1.0   1.0
2   0   4   1  1  1  1  1  1
   0.351076584656   1.0   1.0   1.0   1.0   1.0
2   0   4   1  1  1  1  1  1
   0.174911945670   1.0   1.0   1.0   1.0   1.0
2   0   4   1  1  1  1  1  1
   0.087143916955   1.0   1.0   1.0   1.0   1.0
2   0   4   1  1  1  1  1  1
   0.043416487268   1.0   1.0   1.0   1.0   1.0
```

Boron and Carbon

```
B  LRI-MOLOPT-SR-GTH-MEDIUM
10
2  0  0  1  1
 6.106888146994  1.0
2  0  1  1  1  1
 4.184422872733  1.0  1.0
2  0  2  1  1  1  1
 2.867154982438  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 1.964566666264  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.346115647683  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.922354719774  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.631994903672  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.433041160526  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.296718605830  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.203310768286  1.0  1.0  1.0  1.0  1.0
```

```
C  LRI-MOLOPT-GTH-MEDIUM
15
2  0  0  1  1
12.265249535796  1.0
2  0  0  1  1
 8.393323897961  1.0
2  0  1  1  1  1
 5.743697741369  1.0 1.0
2  0  2  1  1  1  1
 3.930512410253  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 2.689718105443  1.0  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 1.840620950051  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.259568976731  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.861944990411  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.589843970612  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.403640503209  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.276218227104  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.189020944078  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.129350324469  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.088516680105  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.060573506012  1.0  1.0  1.0  1.0  1.0
```

```
C  LRI-MOLOPT-GTH-LARGE
15
2  0  4  1  1  1  1  1  1
12.265249535796  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 8.393323897961  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 5.743697741369  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 3.930512410253  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.689718105443  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.840620950051  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.259568976731  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.861944990411  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.589843970612  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.403640503209  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.276218227104  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.189020944078  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.129350324469  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.088516680105  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.060573506012  1.0  1.0  1.0  1.0  1.0
```

Nitrogen

```
N  LRI-MOLOPT-GTH-MEDIUM
15
2  0  0  1  1
18.085460794596  1.0
2  0  0  1  1
12.261342381151  1.0
2  0  1  1  1  1
 8.312783328846  1.0  1.0
2  0  2  1  1  1 1
 5.635791296275  1.0  1.0  1.0
2  0  3  1  1  1 1  1
 3.820879515162  1.0  1.0  1.0  1.0
2  0  3  1  1  1 1  1
 2.590429542527  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 1.756225284825  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 1.190662475247  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.807229654538  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.547274923593  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.371034240764  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.251548905102  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.170541811795  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.115621690178  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1 1  1  1
 0.078387669856  1.0  1.0  1.0  1.0  1.0
```

```
N  LRI-MOLOPT-GTH-LARGE
15
2  0  4  1  1  1  1  1  1
18.085460794596  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
12.261342381151  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 8.312783328846  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 5.635791296275  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 3.820879515162  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.590429542527  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.756225284825  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.190662475247  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.807229654538  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.547274923593  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.371034240764  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.251548905102  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.170541811795  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.115621690178  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.078387669856  1.0  1.0  1.0  1.0  1.0
```

Oxygen

```
O  LRI-MOLOPT-GTH-MEDIUM
15
2  0  0  1  1
24.031909411024  1.0
2  0  0  1  1
16.167926705922  1.0
2  0  1  1  1  1
10.877281929506  1.0  1.0
2  0  2  1  1  1  1
 7.317899463920  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 4.923256831173  1.0  1.0  1.0  1.0
2  0  3  1  1  1  1  1  1
 3.312215198528  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.228356126354  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.499169204968  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.008594756710  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.678551413605  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.456508441910  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.307124785767  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.206624073890  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.139010297734  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.093521836600  1.0  1.0  1.0  1.0  1.0
```

```
O  LRI-MOLOPT-GTH-LARGE
15
2  0  4  1  1  1  1  1  1
24.031909411024  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
16.167926705922  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
10.877281929506  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 7.317899463920  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 4.923256831173  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 3.312215198528  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.228356126354  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.499169204968  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.008594756710  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.678551413605  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.456508441910  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.307124785767  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.206624073890  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.139010297734  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.093521836600  1.0  1.0  1.0  1.0  1.0
```

Flourine

```
F  LRI-MOLOPT-GTH-MEDIUM
15
2  0  0  1  1
30.051345530254  1.0
2  0  0  1  1
19.938448805906  1.0
2  0  1  1  1  1
13.228750119875  1.0  1.0
2  0  1  1  1  1
 8.777003238198  1.0  1.0
2  0  2  1  1  1  1
 5.823360872741  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 3.863679998041  1.0  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 2.563472100302  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.700810939922  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.128453027836  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.748705341754  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.496750573522  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.329583774194  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.218672045896  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.145084398567  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.096260510214  1.0  1.0  1.0  1.0  1.0
```

```
F  LRI-MOLOPT-GTH-LARGE
15
2  0  4  1  1  1  1  1  1
30.051345530254  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
19.938448805906  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
13.228750119875  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 8.777003238198  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 5.823360872741  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 3.863679998041  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.563472100302  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.700810939922  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.128453027836  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.748705341754  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.496750573522  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.329583774194  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.218672045896  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.145084398567  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.096260510214  1.0  1.0  1.0  1.0  1.0
```

Chlorine

```
Cl LRI-MOLOPT-GTH-MEDIUM
15
2  0  0  1  1
 9.154068388220  1.0
2  0  0  1  1
 6.912319085635  1.0
2  0  1  1  1  1
 5.219554094999  1.0  1.0
2  0  1  1  1  1
 3.941332078728  1.0  1.0
2  0  2  1  1  1  1
 2.976135177849  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 2.247306347170  1.0  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 1.696961164809  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.281390584998  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.967589515524  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.730635515442  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.551709426219  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.416600732575  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.314578947059  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.237541382419  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.179369626890  1.0  1.0  1.0  1.0  1.0
```

```
Cl LRI-MOLOPT-GTH-LARGE
15
2  0  4  1  1  1  1  1  1
 9.154068388220  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 6.912319085635  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 5.219554094999  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 3.941332078728  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.976135177849  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.247306347170  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.696961164809  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.281390584998  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.967589515524  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.730635515442  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.551709426219  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.416600732575  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.314578947059  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.237541382419  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.179369626890  1.0  1.0  1.0  1.0  1.0
```

Bromine

Note that only the short-range version of the Molopt basis sets is available for Br, which are smaller than the standard Molopt basis sets. Thus, the auxiliary basis sets are smaller, too.

```
Br LRI-MOLOPT-SR-GTH-MEDIUM
12
2  0  0  1  1
 3.474785476240  1.0
2  0  1  1  1  1
 2.716635367517  1.0  1.0
2  0  2  1  1  1  1
 2.123903121648  1.0  1.0  1.0
2  0  3  1  1  1  1  1
 1.660496849921  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.298199414321  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.014950265894  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.793502162206  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.620370970464  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.485014609065  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.379191132736  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.296456874614  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.231774086782  1.0  1.0  1.0  1.0  1.0
```

```
Br LRI-MOLOPT-SR-GTH-LARGE
12
2  0  4  1  1  1  1  1  1
 3.474785476240  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.716635367517  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 2.123903121648  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.660496849921  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.298199414321  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 1.014950265894  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.793502162206  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.620370970464  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.485014609065  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.379191132736  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.296456874614  1.0  1.0  1.0  1.0  1.0
2  0  4  1  1  1  1  1  1
 0.231774086782  1.0  1.0  1.0  1.0  1.0
```

B.3. GMTKN24 reactions

Table B.2.: Subsets of the GMTKN24 benchmark suite. Reaction types as given in Supporting Information of Ref [220].

WATER27 subset					
1	$(H_2O)_2$	10	$(H_2O)_{8\ \text{cube}}$ (S_4)	19	$H_3O^+(H_2O)_2$ $(2D)$
2	$(H_2O)_{3\ \text{cyclic}}$	11	$(H_2O)_{20\ \text{dodecahedron}}$	20	$OH^-(H_2O)$
3	$(H_2O)_{4\ \text{cyclic}}$	12	$(H_2O)_{20\ \text{fused cube}}$	21	$OH^-(H_2O)_2$
4	$(H_2O)_{5\ \text{cyclic}}$	13	$(H_2O)_{20\ \text{face-sharing}}$	22	$OH^-(H_2O)_3$
5	$(H_2O)_{6\ \text{prism}}$	14	$(H_2O)_{20\ \text{edge-sharing}}$	23	$OH^-(H_2O)_4(C_4)$
6	$(H_2O)_{6\ \text{cage}}$	15	$H_3O^+(H_2O)$	24	$OH^-(H_2O)_4(C_s)$
7	$(H_2O)_{6\ \text{book}}$	16	$H_3O^+(H_2O)_2$	25	$OH^-(H_2O)_5$
8	$(H_2O)_{6\ \text{cyclic}}$	17	$H_3O^+(H_2O)_3$	26	$OH^-(H_2O)_6$
9	$(H_2O)_{8\ \text{cube}}$ $(D_{2\text{d}})$	18	$H_3O^+(H_2O)_6$ $(3D)$	27	$(H_2O)_{8\ \text{cube}}$ (S_4)-$H_3O^+(H_2O)_6OH^-$

DARC subset	
1	ethene + butadiene
2	ethyne + butadiene
3	ethene + cyclopentadiene
4	ethyne + cyclopentadiene
5	ethene + cyclohexadiene
6	ethyne + cyclohexadiene
7	furan + maleic anhydride (endo-product)
8	furan + maleic anhydride (exo-product)
9	furan + maleimide (endo-product)
10	furan + maleimide (exo-product)
11	cyclopentadiene + maleic anhydride(endo-product)
12	cyclopentadiene + maleic anhydride (exo-product)
13	cyclopentadiene + maleimide (endo-product)
14	cyclopentadiene + maleimide (exo-product)

S22 subset					
1	$(NH_3)_2$	8	$(CH_4)_2$	16	ethene·ethyne
2	$(H_2O)_2$	9	$(C_2H_4)_2$	17	benzene·H_2O
3	formic acid dimer	10	benzene·CH_4	18	benzene·NH_3
4	formamide dimer	11	benzene dimer (C_{2h})	19	benzene·HCN
5	uracil dimer (C_{2h})	12	pyrazine dimer	20	benzene dimer (C_{2v})

Table B.2.: continued

S22 subset					
6	2-pyridoxine·2-aminopyridine	13	uracil dimer (C_2)	21	indole·benzene
7	adenine·thymine	14	indole·benzene	22	phenol dimer
		15	adenine·thymine (stack)		

C. Supplementary information for solid harmonic Gaussian integrals

C.1. The product $\chi_{l,m}(\mathrm{r}_a)r_a^{2n}$ in terms of the STGO

The general formula for generic n of the product $\chi_{l,m}(\mathbf{r}_a)r_a^{2n}$ is given by

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^{2n} &= C_{l,m}(\mathbf{r}_a)\exp\left(-\alpha r_a^2\right)r_a^{2n} \\ &= \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\sum_{j=0}^{n}\binom{n}{j}\frac{(l+j-1)!}{(l-1)!\alpha^j}\exp\left(-\alpha r_a^2\right)r_a^{2(n-j)}\end{aligned} \tag{C.1}$$

where $\mathbf{r}_a = \mathbf{r} - \mathbf{R}_a$. The proof can be found in the next section. For the $(a|b|\tilde{a})$ and $(a|b|\tilde{b})$ integrals required for LRIGPW, the formula has been implemented for $0 \leq n \leq 5$. The implementation for larger n is in fact not necessary. The expressions for the special cases are given by:

Case n=0:

$$\chi_{l,m}(\alpha,\mathbf{r}_a) = \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l} \tag{C.2}$$

Case n=1:

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^2 =& \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^2}{(2\alpha)^l} \\ &+ l\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l\alpha}\end{aligned} \tag{C.3}$$

Case n=2:

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^4 =& \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^4}{(2\alpha)^l} \\ &+ 2l\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^2}{(2\alpha)^l\alpha} \\ &+ l(l+1)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l\alpha^2}\end{aligned} \tag{C.4}$$

Case n=3:

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^6 =& \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^6}{(2\alpha)^l} \\ &+ 3l\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^4}{(2\alpha)^l\alpha} \\ &+ 3l(l+1)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^2}{(2\alpha)^l\alpha^2} \\ &+ l(l+1)(l+2)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l\alpha^3}\end{aligned} \tag{C.5}$$

Case n=4:

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^8 =& \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^8}{(2\alpha)^l} \\ &+ 4l\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^6}{(2\alpha)^l\alpha} \\ &+ 6l(l+1)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^4}{(2\alpha)^l\alpha^2} \\ &+ 4l(l+1)(l+2)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^2}{(2\alpha)^l\alpha^3} \\ &+ l(l+1)(l+2)(l+3)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l\alpha^4}\end{aligned} \tag{C.6}$$

Case n=5:

$$\begin{aligned}\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^{10} =& \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^{10}}{(2\alpha)^l} \\ &+ 5l\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^8}{(2\alpha)^l\alpha} \\ &+ 10l(l+1)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^6}{(2\alpha)^l\alpha^2} \\ &+ 10l(l+1)(l+2)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^4}{(2\alpha)^l\alpha^3} \\ &+ 5l(l+1)(l+2)(l+3)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^2}{(2\alpha)^l\alpha^4} \\ &+ l(l+1)(l+2)(l+3)(l+4)\frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l\alpha^5}\end{aligned} \tag{C.7}$$

C.2. Proof of general formula for $\chi_{l,m}(\mathbf{r}_a)r_a^{2n}$

In the following, we want to prove that

$$\chi_{l,m}(\alpha,\mathbf{r}_a)r_a^{2n} \overset{\text{tbs}}{=} \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\sum_{j=0}^{n}\binom{n}{j}\frac{(l+j-1)!}{(l-1)!\alpha^j}\exp\left(-\alpha r_a^2\right)r_a^{2(n-j)}$$

is valid for all $n \in \mathbb{N}$. The definitions and notations introduced in Section 6.2.1 of Chapter 6 are used. *tbs* indicates that the identity of the left-hand side (lhs) and the right-hand side (rhs) of the equation remains *to be shown*.

Definition 1. The product of a solid harmonic Gaussian function at center $\mathbf{R}_a$ multiplied with the operator r_a^{2n} is defined as

$$\chi_{l,m}(\mathbf{r}_a)r_a^{2n} := C_{l,m}(\mathbf{r}_a)\exp\left(-\alpha r_a^2\right)r_a^{2n}, \tag{C.8}$$

where $C_{l,m}$ is the solid harmonic defined in Equation 6.2 and $n \in \mathbb{N}$.

Definition 2. For all $n \in \mathbb{N}$, let I_n be defined as

$$\mathrm{I}_n := \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\sum_{j=0}^{n}\binom{n}{j}\frac{(l+j-1)!}{(l-1)!\alpha^j}\exp\left(-\alpha r_a^2\right)r_a^{2(n-j)} \tag{C.9}$$

where $C_{l,m}(\nabla_a)$ is the spherical tensor gradient operator (STGO) acting on center $\mathbf{R}_a$. Note that we have dropped the indices writing I_n instead of $\mathrm{I}_n^{l,m,\alpha,\mathbf{r}_a}$.

Definition 3. For all $n \in \mathbb{N}$, let II_n be defined as

$$\begin{aligned}\mathrm{II}_n := \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\exp\left(-\alpha r_a^2\right)r_a^{2n} \\ - \sum_{j=1}^{\min(l,n)}\binom{n}{j}\frac{l!}{(l-j)!}(-\alpha)^{-j}\exp\left(-\alpha r_a^2\right)\left(r_a^2\right)^{n-j}C_{l,m}(\mathbf{r}_a).\end{aligned} \tag{C.10}$$

Note that have we have also dropped the indices $l, m, \alpha, \mathbf{r}_a$ here.

Lemma 1. Let II_n be defined as in Definition 3. Then it holds

$$\mathrm{II}_n = \chi_{l,m}(\alpha,\mathbf{r}_a)r_a^{2n}. \tag{C.11}$$

Proof. The proof is based on Hobson's theorem [244] and Leibniz rule of differentiation [249]. Using Hobson's theorem yields

$$\begin{aligned}C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)r_a^{2n} &= \left[\left(\frac{1}{\mathbf{r}-\mathbf{R}_a}\frac{d}{d\mathbf{R}_a}\right)^l\exp\left(-\alpha r_a^2\right)r_a^{2n}\right]C_{l,m}(\mathbf{r}_a) \\ &= (-2)^l\left[\left(\frac{d}{dr_a^2}\right)^l\exp\left(-\alpha r_a^2\right)r_a^{2n}\right]C_{l,m}(\mathbf{r}_a).\end{aligned} \tag{C.12}$$

For the derivative, the Leibniz rule of differentiation is applied,

$$\begin{aligned}\left(\frac{d}{dr_a^2}\right)^l \exp\left(-\alpha r_a^2\right) r_a^{2n} &= \sum_{j=0}^{\min(l,n)} \binom{l}{j} \left[\left(\frac{d}{dr_a^2}\right)^{l-j} \exp\left(-\alpha r_a^2\right)\right] \left[\left(\frac{d}{dr_a^2}\right)^{j} \left(r_a^2\right)^n\right] \\ &= \sum_{j=0}^{\min(l,n)} \binom{l}{j} (-\alpha)^{l-j} \exp\left(-\alpha r_a^2\right) \frac{n!}{(n-j)!} \left(r_a^2\right)^{n-j} \\ &= \sum_{j=0}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!} (-\alpha)^{l-j} \exp\left(-\alpha r_a^2\right) \left(r_a^2\right)^{n-j}.\end{aligned} \tag{C.13}$$

Inserting the last line in Equation C.12 leads to

$$\begin{aligned}C_{l,m}(\nabla_a) \exp\left(-\alpha r_a^2\right) r_a^{2n} &= (-2)^l \sum_{j=0}^{\min(l,n)} \binom{n}{j} \frac{l!(-\alpha)^{l-j}}{(l-j)!} \exp\left(-\alpha r_a^2\right) \left(r_a^2\right)^{n-j} C_{l,m}(\mathbf{r}_a) \\ &= (-2)^l (-\alpha)^l \exp\left(-\alpha r_a^2\right) \left(r_a^2\right)^n C_{l,m}(\mathbf{r}_a) \\ &\quad + (-2)^l \sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!(-\alpha)^{l-j}}{(l-j)!} \exp\left(-\alpha r_a^2\right) \left(r_a^2\right)^{n-j} C_{l,m}(\mathbf{r}_a)\end{aligned} \tag{C.14}$$

After some rearrangement and employing Definition 1 we get

$$\begin{aligned}\chi_{l,m}(\alpha, \mathbf{r}_a) r_a^{2n} =& \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l} \exp\left(-\alpha r_a^2\right) r_a^{2n} \\ &- \sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!} (-\alpha)^{-j} \exp\left(-\alpha r_a^2\right) \left(r_a^2\right)^{n-j} C_{l,m}(\mathbf{r}_a) \\ =& \mathrm{II}_n.\end{aligned} \tag{C.15}$$

□

Lemma 2. For all $l, m \in \mathbb{N}$, $l \leq m$ it holds that

$$0 = \sum_{j=0}^{l} (-1)^j \binom{l}{j} \frac{(l+m-j-1)!}{(m-j)!}. \tag{C.16}$$

This identity will be used in the proof of Lemma 3.

Proof. For all $l \in \mathbb{N}$ and $x \in \mathbb{R}$ we can employ the binomial formula

$$\left(1 + \frac{1}{x}\right)^l = \sum_{j=0}^{l} \binom{l}{j} \frac{1}{x^j}. \tag{C.17}$$

Multiplication with x^{l+m-1} on both sides yields

$$x^{l+m-1}\left(1+\frac{1}{x}\right)^l = \sum_{j=0}^{l}\binom{l}{j}x^{l+m-j-1}. \tag{C.18}$$

The procedure is now as follows: we take the $(l-1)$-th derivative with respect to x on both sides and set then $x=-1$. We show that the obtained equation corresponds to our hypothesis. The lhs of Equation C.18 is in the following denoted as

$$\mathrm{III}(x) := x^{l+m-1}\left(1+\frac{1}{x}\right)^l \tag{C.19}$$

and the rhs is

$$\mathrm{IV}(x) := \sum_{j=0}^{l}\binom{l}{j}x^{l+m-j-1}. \tag{C.20}$$

Applying the Leibniz rule of differentiation to III yields

$$\left(\frac{d}{dx}\right)^{l-1}\mathrm{III}(x) = \sum_{j=0}^{l-1}\binom{l-1}{j}\left[\left(\frac{d}{dx}\right)^{l-1-j}x^{l+m-1}\right]\left[\left(\frac{d}{dx}\right)^{j}\left(1+\frac{1}{x}\right)^l\right] \tag{C.21}$$

Each of the terms in this sum contains a factor $(1+1/x)^p$ where $p\geq 1, p\in\mathbb{N}$ since we take no more than $l-1$ derivatives. Setting $x=-1$, the factor $(1+1/x)^p$ becomes zero, i.e.

$$\left(\frac{d}{dx}\right)^{l-1}\mathrm{III}(-1) = 0. \tag{C.22}$$

Taking the $(l-1)$-th derivative of IV yields

$$\begin{aligned}\left(\frac{d}{dx}\right)^{l-1}\mathrm{IV}(x) &= \sum_{j=0}^{l}\binom{l}{j}\left(\frac{d}{dx}\right)^{l-1}x^{l+m-j-1}\\ &= \sum_{j=0}^{l}\binom{l}{j}\frac{(l+m-j-1)!}{(m-j)!}x^{m-j}.\end{aligned} \tag{C.23}$$

Notice that $m-j\geq 0$ since $m\geq l$ and $j\leq l$. By inserting $x=-1$, we get

$$\left(\frac{d}{dx}\right)^{l-1}\mathrm{IV}(-1) = \sum_{j=0}^{l}\binom{l}{j}\frac{(l+m-j-1)!}{(m-j)!}(-1)^{m-j}. \tag{C.24}$$

From Equations C.17, C.19 and C.20 we know that the following identity is valid,

$$\left(\frac{d}{dx}\right)^{l-1}\mathrm{III}(-1) = \left(\frac{d}{dx}\right)^{l-1}\mathrm{IV}(-1). \tag{C.25}$$

Inserting Equation C.22 and C.24 and dividing both sides by $(-1)^m$ yields

$$0 = \sum_{j=0}^{l}(-1)^{-j}\binom{l}{j}\frac{(l+m-j-1)!}{(m-j)!}. \tag{C.26}$$

□

Lemma 3. Let I_n and II_n be defined as in Definition 2 and Definition 3. Then it holds that

$$\mathrm{I}_n = \mathrm{II}_n. \tag{C.27}$$

Proof. This is proved by mathematical induction.
1. Basis: we show that $\mathrm{I}_n = \mathrm{II}_n$ for $n = 0$, i.e. $\mathrm{I}_0 = \mathrm{II}_0$. Comparing to Equation C.2, we find that

$$\mathrm{I}_0 = \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l}. \tag{C.28}$$

Since the second term of Equation C.10 yields zero-contribution, we get also

$$\mathrm{II}_0 = \frac{C_{l,m}(\nabla_a)\exp\left(-\alpha r_a^2\right)}{(2\alpha)^l}. \tag{C.29}$$

2. Induction Hypothesis: it holds that $\mathrm{I}_i = \mathrm{II}_i$ for all natural numbers $i < n$.
3. Inductive Step: we have to show that $\mathrm{I}_n = \mathrm{II}_n$ using the induction hypothesis. We have proved in Lemma 1 that

$$\begin{aligned}\mathrm{II}_{n-j} &= \chi_{l,m}(\mathbf{r}_a) r_a^{2(n-j)} \\ &= C_{l,m}(\mathbf{r}_a)\exp(-\alpha r_a^2) r_a^{2(n-j)},\end{aligned} \tag{C.30}$$

where we employed Definition 1 in the last line of Equation C.30. We thus can rewrite II_n as

$$\mathrm{II}_n = \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\exp\left(-\alpha r_a^2\right) r_a^{2n} - \sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j}\mathrm{II}_{n-j}. \tag{C.31}$$

We can now use the induction hypothesis $\mathrm{I}_{n-j} = \mathrm{II}_{n-j}$, since we sum over $j \geq 1$,

$$\mathrm{II}_n = \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\exp\left(-\alpha r_a^2\right) r_a^{2n} - \sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j}\mathrm{I}_{n-j}. \tag{C.32}$$

Employing Definition 2 yields

$$\begin{aligned}\mathrm{II}_n = \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\exp\left(-\alpha r_a^2\right) r_a^{2n} - &\sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j}\frac{C_{l,m}(\nabla_a)}{(2\alpha)^l} \\ &\times \sum_{k=0}^{n-j}\binom{n-j}{k}\frac{(l+k-1)!}{(l-1)!\alpha^k}\exp\left(-\alpha r_a^2\right) r_a^{2(n-j-k)}.\end{aligned} \tag{C.33}$$

In the following, it is shown that Equation C.33 is indeed equivalent to I_n. All terms of I_n with $j > 0$ are in the following denoted as

$$\mathrm{I}'_n := \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l}\sum_{j=1}^{n}\binom{n}{j}\frac{(l+j-1)!}{(l-1)!\alpha^j}\exp\left(-\alpha r_a^2\right) r_a^{2(n-j)} \tag{C.34}$$

and the contributions $j > 0$ for II_n are referred to as

$$\begin{aligned}\mathrm{II}'_n := -\sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j} \frac{C_{l,m}(\nabla_a)}{(2\alpha)^l} \\ \times \sum_{k=0}^{n-j} \binom{n-j}{k} \frac{(l+k-1)!}{(l-1)!\alpha^k} \exp\left(-\alpha r_a^2\right) r_a^{2(n-j-k)}.\end{aligned} \tag{C.35}$$

To prove that $\mathrm{I}_n = \mathrm{II}_n$ it has to be shown that $\mathrm{I}'_n = \mathrm{II}'_n$. Both sides are reduced to get

$$\begin{aligned}&\sum_{j=1}^{n} \binom{n}{j} \frac{(l+j-1)!}{\alpha^j} r_a^{2(n-j)} \\ &\quad \overset{\text{tbs}}{=} -\sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j} \sum_{k=0}^{n-j} \binom{n-j}{k} \frac{(l+k-1)!}{\alpha^k} r_a^{2(n-j-k)},\end{aligned} \tag{C.36}$$

where we denote the rhs by

$$\mathrm{II}''_n := -\sum_{j=1}^{\min(l,n)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j} \sum_{k=0}^{n-j} \binom{n-j}{k} \frac{(l+k-1)!}{\alpha^k} r_a^{2(n-j-k)}. \tag{C.37}$$

In order to remove the sum over $k = 0, ..., n\text{-}j$ in expression II''_n, the δ function is introduced. The exponents of r_a^2 are thereby sorted,

$$\begin{aligned}\mathrm{II}''_n = \sum_{m=1}^{n} r_a^{2(n-m)} \Bigg(-\sum_{j=1}^{\min(l,m)} \binom{n}{j} \\ \times \frac{l!}{(l-j)!}(-\alpha)^{-j} \sum_{k=0}^{n-j} \binom{n-j}{k} \frac{(l+k-1)!}{\alpha^k} \delta_{m,j+k} \Bigg).\end{aligned} \tag{C.38}$$

The range of the newly introduced sum is $m = 1, ..., n$ since for the lower bound of summation we find that $m = j + k \geq 1 + 0 = 1$ and for the upper bound $m = j + k \leq j + (n - j) = n$. For the inner sum over indices j, it must be considered that $k = m - j$ is negative if $j > m$ while the lower bound of the k-sum is in fact $k \geq 0$. Thus, the upper range of the summation of the j-sum has to be changed to $\min(l, n, m)$, which is equivalent to $\min(l, m)$ because $m \leq n$. The summation ranges for the innermost sum are not modified since $k = m - j \leq n - j$. In the next step, the k-sum is eliminated replacing k by $m - j$,

$$\mathrm{II}''_n = \sum_{m=1}^{n} \left(-\sum_{j=1}^{\min(l,m)} \binom{n}{j} \frac{l!}{(l-j)!}(-\alpha)^{-j} \binom{n-j}{m-j} \frac{(l+m-j-1)!}{\alpha^{m-j}} r_a^{2(n-m)} \right). \tag{C.39}$$

Renaming the summation index of the lhs of Equation C.36, we get

$$\mathrm{I}''_n := \sum_{m=1}^{n} \binom{n}{m} \frac{(l+m-1)!}{\alpha^m} r_a^{2(n-m)}. \tag{C.40}$$

It remains to be shown that $\mathrm{I}''_n = \mathrm{II}''_n$ by comparing summand by summand. For each m, the summand on the lhs has to be identical to the summand on the rhs,

$$\binom{n}{m} \frac{(l+m-1)!}{\alpha^m} \overset{\text{tbs}}{=} - \sum_{j=1}^{\min(l,m)} \binom{n}{j} \frac{l!}{(l-j)!} (-\alpha)^{-j} \binom{n-j}{m-j} \frac{(l+m-j-1)!}{\alpha^{m-j}}. \tag{C.41}$$

Reduction of the α parameters and expansion of the binomial coefficients gives

$$\frac{n!(l+m-1)!}{m!(n-m)!} \overset{\text{tbs}}{=} \sum_{j=1}^{\min(l,m)} (-1)^{j+1} \frac{n!}{(n-j)!j!} \frac{l!}{(l-j)!} \frac{(n-j)!(l+m-j-1)!}{(n-j-m+j)!(m-j)!}. \tag{C.42}$$

Further reduction yields

$$\frac{(l+m-1)!}{m!} \overset{\text{tbs}}{=} \sum_{j=1}^{\min(l,m)} (-1)^{j+1} \binom{l}{j} \frac{(l+m-j-1)!}{(m-j)!}. \tag{C.43}$$

The term on the lhs is in fact the negative of the summand $j = 0$ and thus we have

$$0 \overset{\text{tbs}}{=} \sum_{j=0}^{\min(l,m)} (-1)^{j+1} \binom{l}{j} \frac{(l+m-j-1)!}{(m-j)!}. \tag{C.44}$$

It can be verified with common software packages such as Mathematica [253] or its online analogue Wolfram Alpha that the rhs is indeed zero. However, it is easily rationalized dividing Equation C.44 by (-1) and assuming that $\min(l,m) = l$,

$$0 = \sum_{j=0}^{l} (-1)^j \binom{l}{j} \frac{(l+m-j-1)!}{(m-j)!}, \tag{C.45}$$

which corresponds to Lemma 2. In order to show that the rhs of Equation C.44 is also zero for $\min(l,m) = m$, Equation C.43 is reformulated

$$\frac{(l+m-1)!}{l!} \overset{\text{tbs}}{=} \sum_{j=1}^{\min(l,m)} (-1)^{j+1} \binom{m}{j} \frac{(l+m-j-1)!}{(l-j)!} \tag{C.46}$$

The term on the lhs is again the result for $j = 0$ leading to

$$0 = \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{(l+m-j-1)!}{(l-j)!}, \tag{C.47}$$

which also corresponds to Lemma 2. □

Theorem 4. It holds for all $n \in \mathbb{N}$ that

$$\mathrm{I}_n = \chi_{l,m}(\alpha, \mathbf{r}_a) r_a^{2n}. \tag{C.48}$$

Proof. Lemma 1 states that $\mathrm{II}_n = \chi_{l,m}(\alpha, \mathbf{r}_a) r_a^2$ and it holds $\mathrm{I}_n = \mathrm{II}_n$ according to Lemma 3. □

C.3. Integrals $(0_a|r_a^{2m}|0_b)$ and $(0_a|r_b^{2m}|0_b)$ and their derivatives

In the following, the integrals over primitive s functions

$$\begin{aligned}(0_a|r_a^{2m}|0_b) &= \int \chi_{0,0}(\alpha, \mathbf{r}_a) r_a^{2m} \chi_{0,0}(\beta, \mathbf{r}_b) d\mathbf{r} \\ &= \exp(-\zeta_{\alpha\beta} r_{ab}^2) \int \exp(-c r_p^2) r_a^{2m} d\mathbf{r}\end{aligned} \tag{C.49}$$

and

$$\begin{aligned}(0_a|r_b^{2m}|0_b) &= \int \chi_{0,0}(\alpha, \mathbf{r}_a) r_b^{2m} \chi_{0,0}(\beta, \mathbf{r}_b) d\mathbf{r} \\ &= \exp(-\zeta_{\alpha\beta} r_{ab}^2) \int \exp(-c r_p^2) r_b^{2m} d\mathbf{r}\end{aligned} \tag{C.50}$$

as well as their derivatives are explicitly given for $0 \leq m \leq 5, m \in \mathbb{N}$. Note that the Gaussian product rule has been used to summarize the Gaussian functions at $\mathbf{R}_a$ and $\mathbf{R}_b$,

$$\exp(-\alpha r_a^2) \exp(-\beta r_b^2) = \exp(-\zeta_{\alpha\beta} r_{ab}^2) \exp(-c r_p^2) \tag{C.51}$$

yielding a Gaussian function at center $\mathbf{R}_p$,

$$\mathbf{R}_p = \frac{\alpha \mathbf{R}_a + \beta \mathbf{R}_b}{c}, \tag{C.52}$$

where $c = \alpha + \beta$ and $\zeta_{\alpha\beta} = \alpha\beta/c$. Further, the notations $r_{ab}^2 = |\mathbf{R}_a - \mathbf{R}_b|^2$ and $\mathbf{r}_p = \mathbf{r} - \mathbf{R}_p$ are used.

C.3.1. Integral expressions for $(0_a|r_a^{2m}|0_b)$

The evaluation of

$$(0_a|r_a^{2m}|0_b) = \exp(-\zeta_{\alpha\beta} r_{ab}^2) I_{a,m} \tag{C.53}$$

requires the computation of the integral

$$I_{a,m} = \int_{-\infty}^{\infty} \exp(-c r_p^2) r_a^{2m} d\mathbf{r}. \tag{C.54}$$

which has been computed with Mathematica for $0 \leq m \leq 5$. The identity

$$\mathbf{R}_a - \mathbf{R}_p = \frac{\beta}{c}(\mathbf{R}_a - \mathbf{R}_b) \tag{C.55}$$

has been used in order to express the integrals in terms of r_{ab}^2. For $0 \leq m \leq 5$ the results are

Case: m=1

$$I_{a,1} = \frac{\pi^{3/2}\left[3 + \frac{2\beta^2}{c} r_{ab}^2\right]}{2c^{5/2}} \tag{C.56}$$

Case: m=2

$$I_{a,2} = \frac{\pi^{3/2}\left[15 + \frac{20\beta^2}{c} r_{ab}^2 + \frac{4\beta^4}{c^2} r_{ab}^4\right]}{4c^{7/2}} \tag{C.57}$$

Case: m=3

$$I_{a,3} = \frac{\pi^{3/2}\left[105 + \frac{210\beta^2}{c} r_{ab}^2 + \frac{84\beta^4}{c^2} r_{ab}^4 + \frac{8\beta^6}{c^3} r_{ab}^6\right]}{8c^{9/2}} \tag{C.58}$$

Case: m=4

$$I_{a,4} = \frac{\pi^{3/2}\left[945 + \frac{2520\beta^2}{c} r_{ab}^2 + \frac{1512\beta^4}{c^2} r_{ab}^4 + \frac{288\beta^6}{c^3} r_{ab}^6 + \frac{16\beta^8}{c^4} r_{ab}^8\right]}{16c^{11/2}} \tag{C.59}$$

Case: m=5

$$I_{a,5} = \frac{\pi^{3/2}\left[10395 + \frac{34650\beta^2}{c} r_{ab}^2 + \frac{27720\beta^4}{c^2} r_{ab}^4 + \frac{7920\beta^6}{c^3} r_{ab}^6 + \frac{880\beta^8}{c^4} r_{ab}^8 + \frac{32\beta^{10}}{c^5} r_{ab}^{10}\right]}{32c^{13/2}} \tag{C.60}$$

C.3.2. Integral expressions for $(0_a|r_b^{2m}|0_b)$

The calculation of

$$(0_a|r_b^{2m}|0_b) = \exp(-\zeta_{\alpha\beta} r_{ab}^2) I_{b,m} \tag{C.61}$$

is analogous to $(0_a|r_b^{2m}|0_b)$. The integrals $I_{b,m}$,

$$I_{b,m} = \int_{-\infty}^{\infty} \exp(-c r_p^2) r_b^{2m} d\mathbf{r}, \tag{C.62}$$

are computed using here the identity

$$\mathbf{R}_b - \mathbf{R}_p = \frac{\alpha}{c}(\mathbf{R}_b - \mathbf{R}_a) \tag{C.63}$$

to express the integrals in terms of the square distance r_{ab}^2. The only difference to the $I_{a,m}$ integrals is that β^{2j}/c^j is replaced by α^{2j}/c^j, for example

Case: m=1

$$I_{b,1} = \frac{\pi^{3/2}\left[3 + \frac{2\alpha^2}{c} r_{ab}^2\right]}{2c^{5/2}} \tag{C.64}$$

Case: m=2

$$I_{b,2} = \frac{\pi^{3/2}\left[15 + \frac{20\alpha^2}{c} r_{ab}^2 + \frac{4\alpha^4}{c^2} r_{ab}^4\right]}{4c^{7/2}} \tag{C.65}$$

and so on...

C.3.3. Derivatives of $(0_a|r_a^{2m}|0_b)$

The n-th derivative with respect to r_{ab}^2 is obtained by applying the Leibniz rule of differentiation to Equation C.53,

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_a^{2m}|0_b) = \sum_{j=0}^{min(m,n)} \binom{n}{j} (-\zeta_{\alpha\beta})^{n-j} u \left(\frac{d}{dr_{ab}^2}\right)^j v, \tag{C.66}$$

where we have used the abbreviations $u = \exp(-\zeta_{\alpha\beta} r_{ab}^2)$ and $v = I_{a,m}$. For $0 \leq m \leq 5$, the derivatives are

Case: m=1

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_a^2|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} ku \tag{C.67}$$

where

$$k = \frac{\pi^{3/2}\beta^2}{c^{7/2}} \tag{C.68}$$

Case: m=2

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_a^4|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} uv' + n(n-1)(-\zeta_{\alpha\beta})^{n-2} ku \tag{C.69}$$

where

$$v' = \frac{\pi^{3/2}}{4c^{7/2}}\left[\frac{20\beta^2}{c} + \frac{8\beta^4 r_{ab}^2}{c^2}\right] \tag{C.70}$$

and

$$k = \frac{\pi^{3/2}\beta^4}{c^{11/2}} \tag{C.71}$$

Case: m=3

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_a^6|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} uv' + \frac{n(n-1)}{2}(-\zeta_{\alpha\beta})^{n-2} uv'' + n(n-1)(n-2)(-\zeta_{\alpha\beta})^{n-3} ku \tag{C.72}$$

where

$$v' = \frac{\pi^{3/2}}{8c^{9/2}}\left[\frac{210\beta^2}{c} + \frac{168\beta^4 r_{ab}^2}{c^2} + \frac{24\beta^6 r_{ab}^4}{c^3}\right] \tag{C.73}$$

and

$$v'' = \frac{\pi^{3/2}}{8c^{9/2}}\left[\frac{168\beta^4}{c^2} + \frac{48\beta^6 r_{ab}^2}{c^3}\right] \tag{C.74}$$

and

$$k = \frac{\pi^{3/2}\beta^6}{c^{15/2}} \tag{C.75}$$

Case: m=4

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_a^8|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} uv' + \frac{n(n-1)}{2}(-\zeta_{\alpha\beta})^{n-2} uv'' + \frac{n(n-1)(n-2)}{6}(-\zeta_{\alpha\beta})^{n-3} uv'' + n(n-1)(n-2)(n-3)(-\zeta_{\alpha\beta})^{n-4} ku \tag{C.76}$$

where

$$v' = \frac{\pi^{3/2}}{16c^{11/2}}\left[\frac{2520\beta^2}{c} + \frac{3024\beta^4 r_{ab}^2}{c^2} + \frac{864\beta^6 r_{ab}^4}{c^3} + \frac{64\beta^8 r_{ab}^6}{c^4}\right] \tag{C.77}$$

and

$$v'' = \frac{\pi^{3/2}}{16c^{11/2}}\left[\frac{3024\beta^4}{c^2} + \frac{1728\beta^6 r_{ab}^2}{c^3} + \frac{192\beta^8 r_{ab}^4}{c^4}\right] \tag{C.78}$$

and

$$v''' = \frac{\pi^{3/2}}{16c^{11/2}}\left[\frac{1728\beta^6}{c^3} + \frac{384\beta^8 r_{ab}^2}{c^4}\right] \tag{C.79}$$

and

$$k = \frac{\pi^{3/2}\beta^8}{c^{19/2}} \tag{C.80}$$

Case: m=5

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_a^{10}|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} uv' + \frac{n(n-1)}{2}(-\zeta_{\alpha\beta})^{n-2} uv'' + \frac{n(n-1)(n-2)}{6}(-\zeta_{\alpha\beta})^{n-3} uv''' + \frac{n(n-1)(n-2)(n-3)}{24}(-\zeta_{\alpha\beta})^{n-4} uv^{\mathrm{iv}} + n(n-1)(n-2)(n-3)(n-4)(-\zeta_{\alpha\beta})^{n-5} ku \tag{C.81}$$

where

$$v' = \frac{\pi^{3/2}}{32c^{13/2}}\left[\frac{34650\beta^2}{c} + \frac{55440\beta^4 r_{ab}^2}{c^2} + \frac{23760\beta^6 r_{ab}^4}{c^3} + \frac{3520\beta^8 r_{ab}^6}{c^4} + \frac{160\beta^{10} r_{ab}^8}{c^5}\right] \tag{C.82}$$

and

$$v'' = \frac{\pi^{3/2}}{32c^{13/2}}\left[\frac{55440\beta^4}{c^2} + \frac{47520\beta^6 r_{ab}^2}{c^3} + \frac{10560\beta^8 r_{ab}^4}{c^4} + \frac{640\beta^{10} r_{ab}^6}{c^5}\right] \tag{C.83}$$

and

$$v''' = \frac{\pi^{3/2}}{32c^{13/2}}\left[\frac{47520\beta^6}{c^3} + \frac{21120\beta^8 r_{ab}^2}{c^4} + \frac{1920\beta^{10} r_{ab}^4}{c^5}\right] \tag{C.84}$$

and

$$v^{\text{iv}} = \frac{\pi^{3/2}}{32c^{13/2}}\left[\frac{21120\beta^8}{c^4} + \frac{3840\beta^{10} r_{ab}^2}{c^5}\right] \tag{C.85}$$

$$k = \frac{\pi^{3/2}\beta^{10}}{c^{23/2}} \tag{C.86}$$

C.3.4. Derivatives of $(0_a|r_b^{2m}|0_b)$

The derivatives are also taken with respect to r_{ab}^2 for the integrals

$$(0_a|r_b^{2m}|0_b) = \exp(-\zeta_{\alpha\beta} r_{ab}^2) I_{b,m}. \tag{C.87}$$

redefining $v = I_{b,m}$. The expressions for the derivatives are almost the same as for $(0_a|r_b^{2m}|0_b)$ except that the explicit β parameters are replaced by α, for example

Case: m=1

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_b^2|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} ku \tag{C.88}$$

where

$$k = \frac{\pi^{3/2}\alpha^2}{c^{7/2}} \tag{C.89}$$

Case: m=2

$$\left(\frac{d}{dr_{rab}^2}\right)^n (0_a|r_b^4|0_b) = (-\zeta_{\alpha\beta})^n uv + n(-\zeta_{\alpha\beta})^{n-1} uv' + n(n-1)(-\zeta_{\alpha\beta})^{n-2} ku \tag{C.90}$$

where

$$v' = \frac{\pi^{3/2}}{4c^{7/2}}\left[\frac{20\alpha^2}{c} + \frac{8\alpha^4 r_{ab}^2}{c^2}\right] \tag{C.91}$$

and

$$k = \frac{\pi^{3/2}\alpha^4}{c^{11/2}} \tag{C.92}$$

and so on...

Bibliography

[1] A. Warshel, M. Levitt. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

[2] M. J. Field, P. A. Bash, M. Karplus. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

[3] W. Koch, M. C. Holthausen. *A Chemist's Guide to Density Functional Theory.* Wiley-VCH Verlag GmbH, 2nd ed., 2001, p. 19–20.

[4] F. Jensen. *Introduction to Computational Chemistry.* John Wiley & Sons, 2nd ed., 2007, p. 232–267.

[5] P. Hohenberg, W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

[6] L. H. Thomas. The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.

[7] E. Fermi. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Z. Phys.* **1928**, *48*, 73–79.

[8] D. R. Murphy. Sixth-order term of the gradient expansion of the kinetic-energy density functional. *Phys. Rev. A* **1981**, *24*, 1682–1688.

[9] Y. A. Wang, N. Govind, E. A. Carter. Orbital-free kinetic-energy density functionals with a density-dependent kernel. *Phys. Rev. B* **1999**, *60*, 16350–16358.

[10] W. Kohn, L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[11] J. P. Perdew, K. Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2001**, *577*, 1–20.

[12] S. H. Vosko, L. Wilk, M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

[13] J. P. Perdew, Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.

[14] K. Burke. Perspective on density functional theory. *J. Chem. Phys.* **2012**, *136*, 150901.

[15] D. C. Langreth, M. J. Mehl. Beyond the local-density approximation in calculations of ground-state electronic properties. *Phys. Rev. B* **1983**, *28*, 1809–1834.

[16] J. P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.

[17] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **1992**, *46*, 6671–6687.

[18] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[19] C. Lee, W. Yang, R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

[20] J. P. Perdew, K. Burke, M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[21] E. H. Lieb, S. Oxford. Improved lower bound on the indirect Coulomb energy. *Int. J. Quantum Chem.* **1981**, *19*, 427–439.

[22] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[23] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[24] M. Ernzerhof, G. E. Scuseria. Assessment of the Perdew-Burke-Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

[25] M. Fuchs, X. Gonze. Accurate density functionals: Approaches using the adiabatic-connection fluctuation-dissipation theorem. *Phys. Rev. B* **2002**, *65*, 235109.

[26] A. Heßelmann, A. Görling. Random-phase approximation correlation methods for molecules and solids. *Mol. Phys.* **2011**, *109*, 2473–2500.

[27] X. Ren, P. Rinke, C. Joas, M. Scheffler. Random-phase approximation and its applications in computational chemistry and materials science. *J. Mater. Sci.* **2012**, *47*, 7447–7471.

[28] J. Klimeš, A. Michaelides. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, 120901.

[29] S. Zahn, F. Uhlig, J. Thar, C. Spickermann, B. Kirchner. Intermolecular Forces in an Ionic Liquid ([Mmim][Cl]) versus Those in a Typical Salt (NaCl). *Angew. Chem. Int. Ed.* **2008**, *47*, 3639–3641.

[30] A. S. Pensado, M. Brehm, J. Thar, A. P. Seitsonen, B. Kirchner. Effect of Dispersion on the Structure and Dynamics of the Ionic Liquid 1-Ethyl-3-

methylimidazolium Thiocyanate. *ChemPhysChem* **2012**, *13*, 1845–1853.

[31] S. Grimme, W. Hujo, B. Kirchner. Performance of dispersion-corrected density functional theory for the interactions in ionic liquids. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4875–4883.

[32] E. I. Izgorodina, D. Golze, R. Maganti, V. Armel, M. Taige, T. J. S. Schubert, D. R. MacFarlane. Importance of dispersion forces for prediction of thermodynamic and transport properties of some common ionic liquids. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7209–7221.

[33] S. Grimme. Semiempirical GGA-Type Density Functional Constructed With a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[34] S. Grimme, J. Antony, S. Ehrlich, H. Krieg. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

[35] A. Tkatchenko, M. Scheffler. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

[36] A. D. Becke, E. R. Johnson. Exchange-hole dipole moment and the dispersion interaction. *J. Chem. Phys.* **2005**, *122*, 154104.

[37] E. R. Johnson, A. D. Becke. A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

[38] A. D. Becke, E. R. Johnson. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

[39] Y. Andersson, D. C. Langreth, B. I. Lundqvist. Van der Waals Interactions in Density-Functional Theory. *Phys. Rev. Lett.* **1996**, *76*, 102–105.

[40] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

[41] D. C. Langreth, B. I. Lundqvist, S. D. Chakarova-Käck, V. R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P. G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schröder, T. Thonhauser. A density functional for sparse matter. *J. Phys.: Condens. Matter* **2009**, *21*, 084203.

[42] J. Klimeš, D. R. Bowler, A. Michaelides. Chemical accuracy for the van der Waals density functional. *J. Phys.: Condens. Matter* **2010**, *22*, 022201.

[43] M. Franchini, P. H. T. Philipsen, E. van Lenthe, L. Visscher. Accurate Coulomb Potentials for Periodic and Molecular Systems through Density Fitting. *J. Chem. Theory Comput.* **2014**, *10*, 1994–2004.

[44] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

[45] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.

[46] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931–967.

[47] M. Sierka, A. Hogekamp, R. Ahlrichs. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.* **2003**, *118*, 9136–9148.

[48] H. A. Früchtl, R. A. Kendall, R. J. Harrison, K. G. Dyall. An implementation of RI-SCF on parallel computers. *Int. J. Quantum Chem.* **1997**, *64*, 63–69.

[49] F. Weigend. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

[50] R. Polly, H.-J. Werner, F. R. Manby, P. J. Knowles. Fast Hartree-Fock theory using local density fitting approximations. *Mol. Phys.* **2004**, *102*, 2311–2321.

[51] D. E. Bernholdt, R. J. Harrison. Large-scale correlated electronic structure calculations: the RI-MP2 method on parallel computers. *Chem. Phys. Lett.* **1996**, *250*, 477–484.

[52] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143 – 152.

[53] H.-J. Werner, F. R. Manby, P. J. Knowles. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149–8160.

[54] M. Del Ben, J. Hutter, J. VandeVondele. Second-Order Møller-Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach. *J. Chem. Theory Comput.* **2012**, *8*, 4177–4188.

[55] A. P. Rendell, T. J. Lee. Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks. *J. Chem. Phys.* **1994**, *101*, 400–408.

[56] M. Schütz, F. R. Manby. Linear scaling local coupled cluster theory with density fitting. Part I: 4-external integrals. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349–3358.

[57] F. Jensen. *Introduction to Computational Chemistry.* John Wiley & Sons, 2nd ed., 2007, p. 115–118.

[58] T. Bredow, K. Jug. Theory and range of modern semiempirical molecular orbital methods. *Theor. Chem. Acc.* **2005**, *113*, 1–14.

[59] J. A. Pople, D. L. Beveridge, P. A. Dobosh. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. *J.*

Chem. Phys. **1967**, *47*, 2026–2033.

[60] J. A. Pople, D. P. Santry, G. A. Segal. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129–S135.

[61] R. S. Mulliken. Quelques aspects de la théorie des orbitales moléculaires. *J. Chim. Phys.* **1949**, *46*, 497–542.

[62] O. Ayed, E. Bernard, B. Silvi. On the Mulliken approximation of multicentre integrals. *J. Mol. Struct.* **1986**, *135*, 159–168.

[63] K. Rüdenberg. On the Three- and Four-Center Integrals in Molecular Quantum Mechanics. *J. Chem. Phys.* **1951**, *19*, 1433–1434.

[64] P.-O. Löwdin. Approximate Formulas for Many-Center Integrals in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1953**, *21*, 374–375.

[65] M. D. Newton. Self-Consistent Molecular-Orbital Methods. II. Projection of Diatomic Differential Overlap (PDDO). *J. Chem. Phys.* **1969**, *51*, 3917–3926.

[66] F. P. Billingsley, J. E. Bloor. Limited Expansion of Diatomic Overlap (LEDO): A Near-Accurate Approximate Ab Initio LCAO MO Method. I. Theory and Preliminary Investigations. *J. Chem. Phys.* **1971**, *55*, 5178–5190.

[67] H. Sambe, R. H. Felton. A new computational approach to Slater's SCF-Xα equation. *J. Chem. Phys.* **1975**, *62*, 1122–1126.

[68] B. I. Dunlap, J. W. D. Connolly, J. R. Sabin. On some approximations in applications of Xα theory. *J. Chem. Phys.* **1979**, *71*, 3396–3402.

[69] O. Vahtras, J. Almlöf, M. W. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514 – 518.

[70] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* **1997**, *97*, 119–124.

[71] F. Weigend. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

[72] R. E. Stratmann, G. E. Scuseria, M. J. Frisch. Achieving linear scaling in exchange-correlation density functional quadratures. *Chem. Phys. Lett.* **1996**, *257*, 213–223.

[73] O. Treutler, R. Ahlrichs. Efficient molecular numerical integration schemes. *J. Chem. Phys.* **1995**, *102*, 346–354.

[74] D. N. Laikov. Fast evaluation of density functional exchange-correlation terms using the expansion of the electron density in auxiliary basis sets. *Chem. Phys. Lett.* **1997**, *281*, 151–156.

[75] P. Merlot, T. Kjærgaard, T. Helgaker, R. Lindh, F. Aquilante, S. Reine, T. B. Pedersen. Attractive Electron-Electron Interactions within Robust Local Fitting Approximations. *J. Comput. Chem.* **2013**, *34*, 1486–1496.

[76] B. Kirchner, P. J. di Dio, J. Hutter. Real-World Predictions from Ab Initio Molecular Dynamics Simulations. In *Multiscale Molecular Methods in Applied*

Chemistry, B. Kirchner, J. Vrabec, Eds., Volume 307 of *Topics in Current Chemistry*, p. 109–153. Springer Berlin Heidelberg, 2012.

[77] H. M. Senn, W. Thiel. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[78] M. P. Allen, D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, New York, 1992.

[79] H. Lin, D. G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

[80] T. A. Halgren, W. Damm. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.

[81] S. W. Rick, S. J. Stuart. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. In *Reviews in Computational Chemistry, Volume 18*, K. B. Lipkowitz, D. B. Boyd, Eds., p. 89–146. John Wiley & Sons, Inc., 2003.

[82] H. Yu, W. F. van Gunsteren. Accounting for polarization in molecular simulation. *Comput. Phys. Commun.* **2005**, *172*, 69–85.

[83] A. Warshel, M. Kato, A. V. Pisliakov. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

[84] J. Huang, P. E. M. Lopes, B. Roux, A. D. MacKerell Jr. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014**, *5*, 3144–3150.

[85] Y. Shi, P. Ren, M. Schnieders, J.-P. Piquemal. Polarizable Force Fields for Biomolecular Modeling. In *Reviews in Computational Chemistry, Volume 28*, A. L. Parrill, K. B. Lipkowitz, Eds., p. 51–86. John Wiley & Sons, Inc, 2015.

[86] G. Lamoureux, B. Roux. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

[87] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, A. D. MacKerell Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245 – 249.

[88] D. P. Geerke, S. Thiel, W. Thiel, W. F. van Gunsteren. Combined QM/MM Molecular Dynamics Study on a Condensed-Phase S_N2 Reaction at Nitrogen: The Effect of Explicitly Including Solvent Polarization. *J. Chem. Theory Comput.* **2007**, *3*, 1499–1509.

[89] C. J. R. Illingworth, S. R. Gooding, P. J. Winn, G. A. Jones, G. G. Ferenczy, C. A. Reynolds. Classical Polarization in Hybrid QM/MM Methods. *J. Phys. Chem. A* **2006**, *110*, 6487–6497.

[90] D. Golze, M. Iannuzzi, M.-T. Nguyen, D. Passerone, J. Hutter. Simulation of Adsorption Processes at Metallic Interfaces: An Image Charge Augmented

QM/MM Approach. *J. Chem. Theory Comput.* **2013**, *9*, 5086–5097.

[91] M. M. Taddei, T. N. C. Mendes, C. Farina. Subtleties in energy calculations in the image method. *Eur. J. Phys.* **2009**, *30*, 965–972.

[92] D. J. Griffiths. *Introduction to Electrodynamics*. Pearson Education, San Francisco, Calif., 3rd ed., 2008, p. 121–126.

[93] G. Tomba, M. Stengel, W.-D. Schneider, A. Baldereschi, A. De Vita. Supramolecular Self-Assembly Driven by Electrostatic Repulsion: The 1D Aggregation of Rubrene Pentagons on Au(111). *ACS Nano* **2010**, *4*, 7545–7551.

[94] T. Yokoyama, T. Takahashi, K. Shinozaki, M. Okamoto. Quantitative Analysis of Long-Range Interactions between Adsorbed Dipolar Molecules on Cu(111). *Phys. Rev. Lett.* **2007**, *98*, 206102.

[95] E. Wernersson, R. Kjellander. On the effect of image charges and ion-wall dispersion forces on electric double layer interactions. *J. Chem. Phys.* **2006**, *125*, 154702.

[96] L. N. Kantorovich, A. I. Livshits, M. Stoneham. Electrostatic energy calculation for the interpretation of scanning probe microscopy experiments. *J. Phys.: Condens. Matter* **2000**, *12*, 795–814.

[97] F. Bocquet, L. Nony, C. Loppacher. Polarization effects in noncontact atomic force microscopy: A key to model the tip-sample interaction above charged adatoms. *Phys. Rev. B* **2011**, *83*, 035411.

[98] A. Sadeghi, A. Baratoff, S. A. Ghasemi, S. Goedecker, T. Glatzel, S. Kawai, E. Meyer. Multiscale approach for simulations of Kelvin probe force microscopy with atomic resolution. *Phys. Rev. B* **2012**, *86*, 075407.

[99] M. W. Finnis. The interaction of a point charge with an aluminium (111) surface. *Surf. Sci.* **1991**, *241*, 61–72.

[100] M. W. Finnis, R. Kaschner, C. Kruse, J. Furthmüller, M. Scheffler. The interaction of a point charge with a metal surface: theory and calculations for (111), (100) and (110) aluminium surfaces. *J. Phys.: Condens. Matter* **1995**, *7*, 2001–2019.

[101] K. B. Tarmyshov, F. Müller-Plathe. Interface between platinum(111) and liquid isopropanol (2-propanol): A model for molecular dynamics studies. *J. Chem. Phys.* **2007**, *126*, 074702.

[102] F. Iori, S. Corni. Including Image Charge Effects in the Molecular Dynamics Simulations of Molecules on Metal Surfaces. *J. Comput. Chem.* **2008**, *29*, 1656–1666.

[103] F. Iori, R. Di Felice, E. Molinari, S. Corni. GolP: An Atomistic Force-Field to Describe the Interaction of Proteins With Au(111) Surfaces in Water. *J. Comput. Chem.* **2009**, *30*, 1465–1476.

[104] J. I. Siepmann, M. Sprik. Influence of surface topology and electrostatic potential on water/electrode systems. *J. Chem. Phys.* **1995**, *102*, 511–524.

[105] D. T. Limmer, A. P. Willard, P. Madden, D. Chandler. Hydration of metal surfaces can be dynamically heterogeneous and hydrophobic. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 4200–4205.

[106] A. P. Willard, S. K. Reed, P. A. Madden, D. Chandler. Water at an electrochemical interface - a simulation study. *Faraday Discuss.* **2009**, *141*, 423–441.

[107] S. K. Reed, O. J. Lanning, P. A. Madden. Electrochemical interface between an ionic liquid and a model metallic electrode. *J. Chem. Phys.* **2007**, *126*, 084704.

[108] C. Merlet, B. Rotenberg, P. A. Madden, P.-L. Taberna, P. Simon, Y. Gogotsi, M. Salanne. On the molecular origin of supercapacitance in nanoporous carbon electrodes. *Nat. Mater.* **2012**, *11*, 306–310.

[109] J. Vatamanu, L. Cao, O. Borodin, D. Bedrov, G. D. Smith. On the Influence of Surface Topography on the Electric Double Layer Structure and Differential Capacitance of Graphite/Ionic Liquid Interfaces. *J. Phys. Chem. Lett.* **2011**, *2*, 2267–2272.

[110] V. Guallar, F. H. Wallrapp. QM/MM methods: Looking inside heme proteins biochemisty. *Biophys. Chem.* **2010**, *149*, 1–11.

[111] H. Hu, W. Yang. Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

[112] H. M. Senn, W. Thiel. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182–187.

[113] R. A. Friesner, V. Guallar. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzmatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.

[114] D. Fischer, A. Curioni, W. Andreoni. Decanethiols on Gold: The Structure of Self-Assembled Monolayers Unraveled with Computer Simulations. *Langmuir* **2003**, *19*, 3567–3571.

[115] M. L. Sushko, P. V. Sushko, I. V. Abarenkov, A. L. Shluger. QM/MM Method for Metal-Organic Interfaces. *J. Comput. Chem.* **2010**, *31*, 2955–2966.

[116] H. Houjou, Y. Inoue, M. Sakurai. Study of the Opsin Shift of Bacteriorhodopsin: Insight from QM/MM Calculations with Electronic Polarization Effects of the Protein Environment. *J. Phys. Chem. B* **2001**, *105*, 867–879.

[117] M. A. Thompson, G. K. Schenter. Excited States of the Bacteriochlorophyll *b* Dimer of Rhodopseudomonas *viridis*: A QM/MM Study of the Photosynthetic Reaction Center That Includes MM Polarization. *J. Phys. Chem.* **1995**, *99*, 6374–6386.

[118] G. Lippert, J. Hutter, M. Parrinello. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997**, *92*, 477–488.

[119] T. Laino, F. Mohamed, A. Laio, M. Parrinello. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theory Comput.* **2005**, *1*, 1176–

1184.

[120] T. Laino, F. Mohamed, A. Laio, M. Parrinello. An Efficient Linear-Scaling Electrostatic Coupling for Treating Periodic Boundary Conditions in QM/MM Simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378.

[121] The CP2K developers group, CP2K is freely available from: http://www.cp2k.org/ (accessed July, 2015).

[122] J. VandeVondele, J. Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.

[123] S. Goedecker, M. Teter, J. Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.

[124] C. Hartwigsen, S. Goedecker, J. Hutter. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.

[125] M. Krack. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145–152.

[126] S. M. Foiles, M. I. Baskes, M. S. Daw. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys. Rev. B* **1986**, *33*, 7983–7991.

[127] LAMMPS, Large-scale Atomic/Molecular Massively Parallel Simulator, http://lammps.sandia.gov/ (accessed April 22, 2012).

[128] S. Rapino, F. Zerbetto. Modeling the Stability and the Motion of DNA Nucleobases on the Gold Surface. *Langmuir* **2005**, *21*, 2512–2518.

[129] S. F. Boys, F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

[130] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, J. H. van Lenthe. State of the Art in Counterpoise Theory. *Chem. Rev.* **1994**, *94*, 1873–1885.

[131] G. H. Wannier. The Structure of Electronic Excitation Levels in Insulating Crystals. *Phys. Rev.* **1937**, *52*, 191–197.

[132] N. Marzari, D. Vanderbilt. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **1997**, *56*, 12847–12865.

[133] T. S. Chwee, M. B. Sullivan. Adsorption studies of C_6H_6 on Cu (111), Ag (111), and Au (111) within dispersion corrected density functional theory. *J. Chem. Phys.* **2012**, *137*, 134703.

[134] R. F. W. Bader. *Atoms in Molecules: A Quantum Theory.* Clarendon Press, Oxford, 1990, p. 248–351.

[135] W. Tang, E. Sanville, G. Henkelman. A grid-based Bader analysis algorithm without lattice bias. *J. Phys.: Condens. Matter* **2009**, *21*, 084204.

[136] A. Michaelides, V. A. Ranea, P. L. de Andres, D. A. King. General Model for Water Monomer Adsorption on Close-Packed Transition and Noble Metal Surfaces. *Phys. Rev. Lett.* **2003**, *90*, 216102.

[137] H. Ogasawara, B. Brena, D. Nordlund, M. Nyberg, A. Pelmenschikov, L. G. M. Pettersson, A. Nilsson. Structure and Bonding of Water on Pt(111). *Phys. Rev. Lett.* **2002**, *89*, 276102.

[138] A. Hodgson, S. Haq. Water adsorption and the wetting of metal surfaces. *Surf. Sci. Rep.* **2009**, *64*, 381–451.

[139] T. Schiros, K. J. Andersson, L. G. M. Pettersson, A. Nilsson, H. Ogasawara. Chemical bonding of water to metal surfaces studied with core-level spectroscopies. *J. Electron Spectrosc. Relat. Phenom.* **2010**, *177*, 85–98.

[140] S. Nie, P. J. Feibelman, N. C. Bartelt, K. Thürmer. Pentagons and Heptagons in the First Water Layer on Pt(111). *Phys. Rev. Lett.* **2010**, *105*, 026102.

[141] P. J. Feibelman, G. A. Kimmel, R. S. Smith, N. G. Petrik, T. Zubkov, B. D. Kay. A unique vibrational signature of rotated water monolayers on Pt(111): Predicted and observed. *J. Chem. Phys.* **2011**, *134*, 204702.

[142] K. S. McLeary. Physical Properties. In *Water Encyclopedia*, J. H. Lehr, J. Keeley, Eds. John Wiley & Sons, Inc., 2005.

[143] L.-M. Liu, M. Krack, A. Michaelides. Interfacial water: A first principles molecular dynamics study of a nanoscale water film on salt. *J. Chem. Phys.* **2009**, *130*, 234702.

[144] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy. Spatial correlation of dipole fluctuations in liquid water. *Mol. Phys.* **2007**, *105*, 1411–1417.

[145] G. Cicero, J. C. Grossman, E. Schwegler, F. Gygi, G. Galli. Water Confined in Nanotubes and between Graphene Sheets: A First Principle Study. *J. Am. Chem. Soc.* **2008**, *130*, 1871–1878.

[146] T. Todorova, A. P. Seitsonen, J. Hutter, I.-F. W. Kuo, C. J. Mundy. Molecular Dynamics Simulation of Liquid Water: Hybrid Density Functionals. *J. Phys. Chem. B* **2006**, *110*, 3685–3691.

[147] M. Tatarkhanov, D. F. Ogletree, F. Rose, T. Mitsui, E. Fomin, S. Maier, M. Rose, J. I. Cerdá, M. Salmeron. Metal- and Hydrogen-Bonding Competition during Water Adsorption on Pd(111) and Ru(0001). *J. Am. Chem. Soc.* **2009**, *131*, 18425–18434.

[148] A. Michaelides, K. Morgenstern. Ice nanoclusters at hydrophobic metal surfaces. *Nat. Mater.* **2007**, *6*, 597–601.

[149] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, U. Rothlisberger. Structure and Dynamics of Liquid Water from ab Initio Molecular Dynamics - Comparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections. *J. Chem. Theory Comput.* **2012**, *8*, 3902–3910.

[150] A. Luzar, D. Chandler. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys. Rev. Lett.* **1996**, *76*, 928–931.

[151] A. Luzar, D. Chandler. Hydrogen-bond kinetics in liquid water. *Nature* **1996**, *379*, 55–57.

[152] P. Mark, L. Nilsson. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.

[153] S. Chatterjee, P. G. Debenedetti, F. H. Stillinger, R. M. Lynden-Bell. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. *J. Chem. Phys.* **2008**, *128*, 124511.

[154] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy. Vapor-liquid equilibria of water from first principles: comparison of density functionals and basis sets. *Mol. Phys.* **2006**, *104*, 3619–3626.

[155] D. Golze, J. Hutter, M. Iannuzzi. Wetting of water on hexagonal boron nitride@Rh(111): a QM/MM model based on atomic charges derived for nanostructured substrates. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14307–14316.

[156] M. A. Henderson. The interaction of water with solid surfaces: fundamental aspects revisited. *Surf. Sci. Rep.* **2002**, *46*, 1–308.

[157] P. A. Thiel, T. E. Madey. The interaction of water with solid surfaces: Fundamental aspects. *Surf. Sci. Rep.* **1987**, *7*, 211–385.

[158] M. Corso, W. Auwärter, M. Muntwiler, A. Tamai, T. Greber, J. Osterwalder. Boron Nitride Nanomesh. *Science* **2004**, *303*, 217–220.

[159] R. Laskowski, P. Blaha, K. Schwarz. Bonding of hexagonal BN to transition metal surfaces: An *ab initio* density-functional theory study. *Phys. Rev. B* **2008**, *78*, 045409.

[160] J. Gómez Díaz, Y. Ding, R. Koitz, A. P. Seitsonen, M. Iannuzzi, J. Hutter. Hexagonal boron nitride on transition metal surfaces. *Theor. Chem. Acc.* **2013**, *132*, 1350.

[161] S. Berner, M. Corso, R. Widmer, O. Groening, R. Laskowski, P. Blaha, K. Schwarz, A. Goriachko, H. Over, S. Gsell, M. Schreck, H. Sachdev, T. Greber, J. Osterwalder. Boron Nitride Nanomesh: Functionality from a Corrugated Monolayer. *Angew. Chem. Int. Ed.* **2007**, *46*, 5115–5119.

[162] A. B. Preobrajenski, A. S. Vinogradov, N. Mårtensson. Monolayer of h-BN chemisorbed on Cu(111) and Ni(111): The role of the transition metal 3d states. *Surf. Sci.* **2005**, *582*, 21–30.

[163] A. B. Preobrajenski, S. A. Krasnikov, A. S. Vinogradov, M. L. Ng, T. Käämbre, A. A. Cafolla, N. Mårtensson. Adsorption-induced gap states of *h*-BN on metal surfaces. *Phys. Rev. B* **2008**, *77*, 085421.

[164] I. Shimoyama, Y. Baba, T. Sekiguchi, K. G. Nath. A theoretical interpretation of near edge X-ray absorption fine structure of hexagonal boron nitride monolayer on Ni(111). *J. Electron Spectrosc. Relat. Phenom.* **2009**, *175*, 6–13.

[165] S. Joshi, D. Ecija, R. Koitz, M. Iannuzzi, A. P. Seitsonen, J. Hutter, H. Sachdev, S. Vijayaraghavan, F. Bischoff, K. Seufert, J. V. Barth, W. Auwärter. Boron Nitride on Cu(111): An Electronically Corrugated Monolayer. *Nano Lett.* **2012**, *12*, 5821–5828.

[166] S. Joshi, F. Bischoff, R. Koitz, D. Ecija, K. Seufert, A. P. Seitsonen, J. Hutter, K. Diller, J. I. Urgel, H. Sachdev, J. V. Barth, W. Auwärter. Control of Molecular Organization and Energy Level Alignment by an Electronically Nanopatterned Boron Nitride Template. *ACS Nano* **2014**, *8*, 430–442.

[167] F. Orlando, R. Larciprete, P. Lacovig, I. Boscarato, A. Baraldi, S. Lizzit. Epitaxial Growth of Hexagonal Boron Nitride on Ir(111). *J. Phys. Chem. C* **2012**, *116*, 157–164.

[168] F. Schulz, R. Drost, S. K. Hämäläinen, P. Liljeroth. Templated Self-Assembly and Local Doping of Molecules on Epitaxial Hexagonal Boron Nitride. *ACS Nano* **2013**, *7*, 11121–11128.

[169] A. B. Preobrajenski, A. S. Vinogradov, M. L. Ng, E. Ćavar, R. Westerström, A. Mikkelsen, E. Lundgren, N. Mårtensson. Influence of chemical interaction at the lattice-mismatched *h*-BN/Rh(111) and *h*-BN/Pt(111) interfaces on the overlayer morphology. *Phys. Rev. B* **2007**, *75*, 245412.

[170] M. Morscher, M. Corso, T. Greber, J. Osterwalder. Formation of single layer h-BN on Pd(111). *Surf. Sci.* **2006**, *600*, 3280 – 3284.

[171] F. Müller, S. Hüfner, H. Sachdev, R. Laskowski, P. Blaha, K. Schwarz. Epitaxial growth of hexagonal boron nitride on Ag(111). *Phys. Rev. B* **2010**, *82*, 113406.

[172] T. Brugger, S. Günther, B. Wang, J. H. Dil, M.-L. Bocquet, J. Osterwalder, J. Wintterlin, T. Greber. Comparison of electronic structure and template function of single-layer graphene and a hexagonal boron nitride nanomesh on Ru(0001). *Phys. Rev. B* **2009**, *79*, 045407.

[173] A. Goriachko, Y. He, M. Knapp, H. Over, M. Corso, T. Brugger, S. Berner, J. Osterwalder, T. Greber. Self-Assembly of a Hexagonal Boron Nitride Nanomesh on Ru(0001). *Langmuir* **2007**, *23*, 2928–2931.

[174] R. Laskowski, P. Blaha, T. Gallauner, K. Schwarz. Single-Layer Model of the Hexagonal Boron Nitride Nanomesh on the Rh(111) Surface. *Phys. Rev. Lett.* **2007**, *98*, 106802.

[175] R. Laskowski, P. Blaha. Unraveling the structure of the h-BN/Rh(111) nanomesh with *ab initio* calculations. *J. Phys.: Condens. Matter* **2008**, *20*, 064207.

[176] H. Dil, J. Lobo-Checa, R. Laskowski, P. Blaha, S. Berner, J. Osterwalder, T. Greber. Surface Trapping of Atoms and Molecules with Dipole Rings. *Science* **2008**, *319*, 1824–1826.

[177] M. Iannuzzi, F. Tran, R. Widmer, T. Dienel, K. Radican, Y. Ding, J. Hutter, O. Gröning. Site-selective adsorption of phthalocyanine on h-BN/Rh(111)

nanomesh. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12374–12384.

[178] P. Bacle, A. P. Seitsonen, M. Iannuzzi, J. Hutter. Chemical Reactions on Metal-supported Hexagonal Boron Nitride Investigated with Density Functional Theory. *Chimia* **2014**, *68*, 596–601.

[179] R. Widmer, D. Passerone, T. Mattle, H. Sachdev, O. Gröning. Probing the selectivity of a nanostructured surface by xenon adsorption. *Nanoscale* **2010**, *2*, 502–508.

[180] H. Ma, T. Brugger, S. Berner, Y. Ding, M. Iannuzzi, J. Hutter, J. Osterwalder, T. Greber. Nano-ice on Boron Nitride Nanomesh: Accessing Proton Disorder. *ChemPhysChem* **2010**, *11*, 399–403.

[181] H. Ma, Y. Ding, M. Iannuzzi, T. Brugger, S. Berner, J. Hutter, J. Osterwalder, T. Greber. Chiral Distortion of Confined Ice Oligomers (n = 5,6). *Langmuir* **2012**, *28*, 15246–15250.

[182] Y. Ding, M. Iannuzzi, J. Hutter. Investigation of Boron Nitride Nanomesh Interacting with Water. *J. Phys. Chem. C* **2011**, *115*, 13685–13692.

[183] Y. Ding, M. Iannuzzi, J. Hutter. Investigation of *h*-BN/Rh(111) Nanomesh Interacting with Water and Atomic Hydrogen. *Chimia* **2011**, *65*, 256–259.

[184] Y. Ding, M. Iannuzzi, J. Hutter. Nano-ice models for the water aggregates observed on the h-BN/Rh(111) nanomesh. *J. Phys.: Condens. Matter* **2012**, *24*, 445002.

[185] O. Bunk, M. Corso, D. Martoccia, R. Herger, P. R. Willmott, B. D. Patterson, J. Osterwalder, J. F. van der Veen, T. Greber. Surface X-ray diffraction study of boron-nitride nanomesh in air. *Surf. Sci.* **2007**, *601*, L7–L10.

[186] R. Widmer, S. Berner, O. Gröning, T. Brugger, J. Osterwalder, T. Greber. Electrolytic in situ STM investigation of h-BN-Nanomesh. *Electrochem. Commun.* **2007**, *9*, 2484–2488.

[187] T. Brugger, H. Ma, M. Iannuzzi, S. Berner, A. Winkler, J. Hutter, J. Osterwalder, T. Greber. Nanotexture Switching of Single-Layer Hexagonal Boron Nitride on Rhodium by Intercalation of Hydrogen Atoms. *Angew. Chem. Int. Ed.* **2010**, *49*, 6120–6124.

[188] C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[189] C. Campañá, B. Mussard, T. K. Woo. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *J. Chem. Theory Comput.* **2009**, *5*, 2866–2878.

[190] D.-L. Chen, A. C. Stern, B. Space, J. K. Johnson. Atomic Charges Derived from Electrostatic Potentials for Molecular and Periodic Systems. *J. Phys. Chem. A* **2010**, *114*, 10225–10233.

[191] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele. CP2K: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci* **2014**, *4*, 15–25.

[192] Y. Zhang, W. Yang. Comment on "Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.* **1998**, *80*, 890.

[193] M. Weinert, J. W. Davenport. Fractional occupations and density-functional energies and forces. *Phys. Rev. B* **1992**, *45*, 13709–13712.

[194] P. Bendt, A. Zunger. New approach for solving the density-functional self-consistent-field problem. *Phys. Rev. B* **1982**, *26*, 3114–3137.

[195] J. W. Kang, H. J. Hwang. Comparison of C_{60} encapsulations into carbon and boron nitride nanotubes. *J. Phys.: Condens. Matter* **2004**, *16*, 3901–3908.

[196] C. Y. Won, N. R. Aluru. Water Permeation through a Subnanometer Boron Nitride Nanotube. *J. Am. Chem. Soc.* **2007**, *129*, 2748–2749.

[197] Y. Wu, H. L. Tepper, G. A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006**, *124*, 024503.

[198] T. Young. An Essay on the Cohesion of Fluids. *Phil. Trans. R. Soc. Lond.* **1805**, *95*, 65–87.

[199] P. G. de Gennes. Wetting: statics and dynamics. *Rev. Mod. Phys.* **1985**, *57*, 827–863.

[200] Y. Yuan, T. R. Lee. *Surface Science Techniques*, Chapter: Contact Angle and Wetting Properties. Springer, 2013, p. 3–34.

[201] B. Lange, R. Posner, K. Pohl, C. Thierfelder, G. Grundmeier, S. Blankenburg, W. G. Schmidt. Water adsorption on hydrogenated Si(111) surfaces. *Surf. Sci.* **2009**, *603*, 60–64.

[202] P. E. Blöchl. Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *J. Chem. Phys.* **1995**, *103*, 7422–7428.

[203] M. V. Fernández-Serra, E. Artacho. Network equilibration and first-principles liquid water. *J. Chem. Phys.* **2004**, *121*, 11136–11144.

[204] D. Prendergast, G. Galli. X-Ray Absorption Spectra of Water from First Principles Calculations. *Phys. Rev. Lett.* **2006**, *96*, 215502.

[205] J. L. Whitten. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.

[206] B. I. Dunlap, J. W. D. Connolly, J. R. Sabin. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993–4999.

[207] E. J. Baerends, D. E. Ellis, P. Ros. Self-consistent molecular Hartree-Fock-Slater calculations I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41–51.

[208] A. Sodt, M. Head-Gordon. Hartree-Fock exchange computed using the atomic resolution of the identity approximation. *J. Chem. Phys.* **2008**, *128*, 104106.

[209] S. F. Manzer, E. Epifanovsky, M. Head-Gordon. Efficient Implementation of the Pair Atomic Resolution of the Identity Approximation for Exact Exchange for

Hybrid and Range-Separated Density Functionals. *J. Chem. Theory Comput.* **2015**, *11*, 518–527.

[210] D. Mejía-Rodríguez, A. M. Köster. Robust and efficient variational fitting of Fock exchange. *J. Chem. Phys.* **2014**, *141*, 124114.

[211] C. Fonseca Guerra, J. G. Snijders, G. te Velde, E. J. Baerends. Towards an order-N DFT method. *Theor. Chem. Acc.* **1998**, *99*, 391–403.

[212] Y. Jung, A. Sodt, P. M. W. Gill, M. Head-Gordon. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6692–6697.

[213] S. Reine, E. Tellgren, A. Krapp, T. Kjærgaard, T. Helgaker, B. Jansik, S. Høst, P. Salek. Variational and robust density fitting of four-center two-electron integrals in local metrics. *J. Chem. Phys.* **2008**, *129*, 104101.

[214] B. I. Dunlap. Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *J. Mol. Struct. (Theochem)* **2000**, *529*, 37–40.

[215] M. J. D. Powell. The NEWUOA software for unconstrained optimization without derivatives. In *Large-Scale Nonlinear Optimization*, G. Di Pillo, M. Roma, Eds., Volume 83 of *Nonconvex Optimization and Its Applications*, p. 255–297. Springer US, 2006.

[216] S. Obara, A. Saika. Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. *J. Chem. Phys.* **1986**, *84*, 3963–3974.

[217] G. H. Golub, C. F. Van Loan. *Matrix computation*. The Johns Hopkins Unversity Press, Baltimore, 2nd ed., 1989, p. 243.

[218] J. VandeVondele, J. Hutter. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365–4369.

[219] A. D. Boese, G. Jansen, M. Torheyden, S. Höfener, W. Klopper. Effects of counterpoise correction and basis set extrapolation on the MP2 geometries of hydrogen bonded dimers of ammonia, water, and hydrogen fluoride. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1230–1238.

[220] L. Goerigk, S. Grimme. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions - Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107–126.

[221] C. G. Salzmann, P. G. Radaelli, E. Mayer, J. L. Finney. Ice XV: A New Thermodynamically Stable Phase of Ice. *Phys. Rev. Lett.* **2009**, *103*, 105701.

[222] C. G. Salzmann, P. G. Radaelli, B. Slater, J. L. Finney. The polymorphism of ice: five unresolved questions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18468–18480.

[223] K. D. Nanda, G. J. O. Beran. What Governs the Proton Ordering in Ice XV? *J. Phys. Chem. Lett.* **2013**, *4*, 3165–3169.

[224] M. Del Ben, J. VandeVondele, B. Slater. Periodic MP2, RPA, and Boundary Condition Assessment of Hydrogen Ordering in Ice XV. *J. Phys. Chem. Lett.* **2014**, *5*, 4122–4128.

[225] M. V. Berry. Quantal Phase Factors Accompanying Adiabatic Changes. *Proc. Roy Soc. of London A* **1984**, *392*, 45–57.

[226] F. H. Allen. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallgr. B* **2002**, *58*, 380–388.

[227] L. Maschio, D. Usvyat, M. Schütz, B. Civalleri. Periodic local Møller-Plesset second order perturbation theory method applied to molecular crystals: Study of solid NH_3 and CO_2 using extended basis sets. *J. Chem. Phys.* **2010**, *132*, 134706.

[228] K. N. Rankin, R. J. Boyd. A Density Functional Theory Study of the Dimers of HX (X=F, Cl, and Br). *J. Comput. Chem.* **2001**, *22*, 1590–1597.

[229] P. Zielke, M. A. Suhm. Raman jet spectroscopy of formic acid dimers: low frequency vibrational dynamics and beyond. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4528–4534.

[230] M. Del Ben, J. Hutter, J. VandeVondele. Forces and stress in second order Møller-Plesset perturbation theory for condensed phase systems within the resolution-of-identity Gaussian and plane waves approach. *J. Chem. Phys.* **2015**, *143*, 102803.

[231] L. Maschio, B. Civalleri, P. Ugliengo, A. Gavezzotti. Intermolecular Interaction Energies in Molecular Crystals: Comparison and Agreement of Localized Møller-Plesset 2, Dispersion-Corrected Density Functional, and Classical Empirical Two-Body Calculations. *J. Phys. Chem. A* **2011**, *115*, 11179–11186.

[232] T. J. Giese, D. M. York. Contracted auxiliary Gaussian basis integral and derivative evaluation. *J. Chem. Phys.* **2008**, *128*, 064104.

[233] L. E. McMurchie, E. R. Davidson. One- and two-electron integrals over cartesian gaussian functions. *J. Comput. Phys.* **1978**, *26*, 218–231.

[234] T. P. Hamilton, H. F. Schaefer III. New variations in two-electron integral evaluation in the context of direct SCF procedures. *Chem. Phys.* **1991**, *150*, 163–171.

[235] R. Lindh, U. Ryu, B. Liu. The reduced multiplication scheme of the Rys quadrature and new recurrence relations for auxiliary function based two-electron integral evaluation. *J. Chem. Phys.* **1991**, *95*, 5889–5897.

[236] B. I. Dunlap. Three-center Gaussian-type-orbital integral evaluation using solid spherical harmonics. *Phys. Rev. A* **1990**, *42*, 1127–1137.

[237] B. I. Dunlap. Direct quantum chemical integral evaluation. *Int. J. Quantum Chem.* **2001**, *81*, 373–383.

[238] B. I. Dunlap. Angular momentum in solid-harmonic-Gaussian integral evaluation. *J. Chem. Phys.* **2003**, *118*, 1036–1043.

[239] A. Hu, B. I Dunlap. Three-center molecular integrals and derivatives using solid harmonic Gaussian orbital and Kohn-Sham potential basis sets. *Can. J. Chem.* **2013**, *91*, 907–915.

[240] J. Kuang, C. D. Lin. Molecular integrals over spherical Gaussian-type orbitals: I. *J. Phys. B: At. Mol. Opt. Phys.* **1997**, *30*, 2529–2548.

[241] J. Kuang, C. D. Lin. Molecular integrals over spherical Gaussian-type orbitals: II. Modified with plane-wave phase factors. *J. Phys. B: At. Mol. Opt. Phys.* **1997**, *30*, 2549–2567.

[242] E. J. Weniger, E. O. Steinborn. A simple derivation of the addition theorems of the irregular solid harmonics, the Helmholtz harmonics, and the modified Helmholtz harmonics. *J. Math. Phys.* **1985**, *26*, 664–670.

[243] E. J. Weniger. The Spherical Tensor Gradient Operator. *arXiv:math-ph/0505018v1* **2005**.

[244] E. W. Hobson. On a Theorem in Differentiation, and its application to Spherical Harmonics. *Proc. London Math. Soc.* **1892**, *s1-24*, 55–67.

[245] M. A. Watson, P. Sałek, P. Macak, T. Helgaker. Linear-scaling formation of Kohn-Sham Hamiltonian: Application to the calculation of excitation energies and polarizabilities of large molecular systems. *J. Chem. Phys.* **2004**, *121*, 2915–2931.

[246] T. Helgaker, P. Jørgensen, J. Olsen. *Molecular Electron-Structure Theory.* John Wiley & Sons, 2012, p. 412–416.

[247] H. B. Schlegel, M. J. Frisch. Transformation between Cartesian and pure spherical harmonic Gaussians. *Int. J. Quantum Chem.* **1995**, *54*, 83–87.

[248] A. Hu, M. Staufer, U. Birkenheuer, V. Igoshine, N. Rösch. Analytical evaluation of pseudopotential matrix elements with Gaussian-type solid harmonics of arbitrary angular momentum. *Int. J. Quantum Chem.* **2000**, *79*, 209–221.

[249] I. N. Bronstein, K. A. Semendjajew, G. Musiol, M. Mühlig. *Taschenbuch der Mathematik.* Verlag Europa-Lehrmittel, Edition Harri Deutsch, 9th ed., 2013, p. 451–452.

[250] J. M. Pérez-Jordá, W. Yang. A concise redefinition of the solid spherical harmonics and its use in fast multipole methods. *J. Chem. Phys.* **1996**, *104*, 8003–8006.

[251] G. Lippert, J. Hutter, M. Parrinello. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **1999**, *103*, 124–140.

[252] J. VandeVondele, M. Iannuzzi, J. Hutter. Large Scale Condensed Matter Calculations using the Gaussian and Augmented Plane Waves Method. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, M. Ferrario, G. Ciccotti, K. Binder, Eds., Lecture Notes in Physics 703, p. 287–314. Springer Berlin Heidelberg, 2006.

[253] Wolfram Research; Inc. *Mathematica, Version 8.0.* Champaign, Illinois, 2010.